UNIVERSITI TEKNOLOGI MARA

MATHEMATICS AND STATISTICS

UNDERGRADUATE RESEARCH PROCEEDINGS 2025



UiTM CAWANGAN NEGERI SEMBILAN

MATHEMATICS AND STATISTICS

UNDERGRADUATE RESEARCH PROCEEDINGS 2025

All articles in this journal may not be altered, copied, distributed, retransmitted, broadcast, displayed, published, licensed, transferred, or sold in any way without written permission.

DISCLAIMER: All articles published in this journal represent the knowledge and personal opinions of the article authors. Readers are advised to seek professional advice before following any information in this journal. The university, publisher, and editorial board will not be responsible or liable for any direct or indirect losses resulting from the information presented.

PUBLISHED BY:

Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, Persiaran Seremban Tiga/1, Seremban 3, 70300 Seremban,
NEGERI SEMBILAN, MALAYSIA.

Tel: +606 634 2000, Fax: +606 633 5813
eISBN: 9786299595328
2025 COPYRIGHT

UNIVERSITI TEKNOLOGI MARA

MATHEMATICS AND STATISTICS

UNDERGRADUATE RESEARCH PROCEEDINGS 2025

We express our gratitude to Allah SWT for the successful publication of the Mathematics and Statistics Undergraduate Research Proceedings 2025. This proceeding is published by the Faculty of Computer and Mathematical Sciences (FSKM) at UiTM Seremban Campus. The publication aims to showcase the research contributions of students and lecturers involved in the Final Year Projects of the Bachelor's Degree Programme at FSKM, UiTM Seremban. These projects represent significant academic efforts, yet many have faced challenges in being published in academic conference proceedings or journals due to factors such as scientific rigor and writing quality. We hope that this publication will serve as a valuable platform for disseminating research findings from FSKM Seremban and provide greater visibility for the outstanding work carried out by our students and lecturers. On behalf of the editorial team, I would like to extend my sincere appreciation and commendation to the publishing team for their commitment and hard work in bringing this proceeding to fruition. Thank you.

Editor in Chief

EDITORIAL

ADVISORY BOARD

Prof. Dr. Yamin Yasin
Rector, UiTMCNS

Prof. Madya Dr. Noorlis Ahmad
Deputy Rector (Academic Affais), UiTMCNS

EDITOR IN CHIEF

Dr. Mat Salim Selamat
Senior Lecturer, FSKM, UiTMCNS

MANAGING EDITOR

Dr. Nur Amalina Shafie

GRAPHICS EDITOR

Dr. Intan Syaherra Ramli
Muhamad Luqman Sapini

EDITORIAL BOARD

Dr. Noorehan Awang
Dr. Nurul Liyana Abdul Aziz
Dr. Rossidah Wan Abdul Aziz
Dr. Siti Hidayah Muhad Saleh
Dr. Siti Sarah Januri
Dr. Zahari Md Rodzi
Dr. Zati Aqmar Zaharudin
Haslinda Ab Malek
Isnewati Ab Malek
Noorezatty Mohd Yusop
Nor Faradilah Mahad
Noreha Mohamed Yusof
Norliza Muhamad Yusof

LANGUAGE EDITOR

AP Dr. Alice Shanthi
AP Dr. Angeline Ranjethamoney Vijayarajoo
Dr. Tuan Sarifah Aini Syed Ahmad
Dr. Wan Zumusni Wan Mustapha
Dr. Irma Ahmad
Dr. Nurul Kamalia Yusuf
Nur Amirah Nabihah Zainal Abidin
Fatin Aliyah Hassan
Noor Shahariah Binti Saleh
Ruhil Nadiah Abu Bakar
Sheela Paramasivam
Asmahanim Haji Mohamad Yusuf
Nurul Syafawani Halim
Mohd Ikram Mohd Nor Rizan
Sharina Salmi Azmi
Tengku Farah Petri Tengku Mahmood
Norashikin Mohd Mokhtar
Mohd Nur Fitri Bin Mohd Salim

MATHEMATICS AND STATISTICS UNDERGRADUATE RESEARCH PROCEEDINGS 2025

Fakulti Sains Komputer dan Matematik
Universiti Teknologi MARA Cawangan Negeri Sembilan

Contents

The Importance of Academic Writing to Undergraduate Mathematics and Statistics Students: A Case for Student Final Year Project Proceedings
Mat Salim Selamat **1**

Mortality Modelling And Forecasting: An Application Of The Lee-Carter Model On The Japanese Data
Sherlinda Dinda Lim & Nurul Aityqah Yaacob **5**

Air Quality Assessment In Klang Using Principal Component Analysis
Nur Afrina Syamimi Norazman & Isnewati Ab Malek **17**

Analyzing Attrition Among The Married Employees
Amanda Balqish Alri Mohammad Haelmi Alri & Noreha Mohamed Yusof **23**

Analysis Of Waiting Lines And Service Facilities: A Study Using M/M/1 And M/D/1 Queuing Models
Stephanie Morissa Anak Lister Neddi & Najir Tokachil **30**

Classification Symptoms Of Mental Health Problems Among Students By Using Fuzzy Inference System
Muhammad Faiq Zhafran bin Abdul Gani, Muhammad Hani Akmal bin Hefnee, Mustafa Nabil bin Abd Malek, Siti Nazifah Zainol Abidin & Hazwa Hanim binti Mohamed Hamzah **37**

Coefficient Bound Of A Generalized Class Of Close-To-Convex Functions Using Herglotz Representation Theorem
Anisa Nur Syahira Lokman & Abdullah Yahya **51**

Factors Affecting Rental Prices Of Residential Apartments In Kuala Lumpur
Nur Ainaa Azrul, Siti Aishah Mohd Shafie, Nornadiah Mohd Razali & Az'lina Abdul Hadi **59**

Forecasting Stock Price In Healthcare Sector By Using Geometric Brownian Motion Model
Muhammad Afiq Mukhriz bin Shahruzain, Muhammad Nur Hazwan bin Zulfemi, Puteri Nurul Atiqah binti Muhamad Nor, Siti Nazifah Zainol Abidin & Aslina binti Omar **69**

Identifying Factors Shaping Urban Poverty In Malaysia: A Household Perspective
Elis Eliza binti Rosli, Siti Aishah binti Mohd Shafie, Az'lina Abdul Hadi & Nornadiah Mohd Razali **79**

Chaos In The Waters: Investigating Non-Linear Dynamics Of Rainfall In River Basin
Muhammad Mirza Irfan Bin Mohd Nor & Muhamad Luqman Sapini **90**

Application Of Dijkstra's Algorithm To Find The Minimum Travel Time For Public Transport In A City
Nurin Aina Sofhia Binti Zuraidi, Noor Hidayah Mohd Zaki & Noraimi Azlin Mohd Nordin **102**

Shortest Route Optimization For Emergency Cases Using Integer Linear Programming
Nurul Liyana Abdul Aziz & Aisyah Fariha Norzaki **120**

Financial Performance Evaluation Of Companies Using Topsis Approach: A Case Study Of Properties Sector In Malaysia
Ariesa Sofea Mohd Yusof & Farah Waheeda Azhar **127**

Mud Crab Weight Prediction Using Multiple Linear Regression Model
Maisarah Saqinah binti Mohd Ridzwan & Yusrina Andu **134**

Evaluating And Enhancing Students' Level Of Financial Management: A Case Study
Nora Mohd Basir & Nurafrina Batrisyia Johari **139**

Numerical Solution Of Solving Mhd Flow Over A Shrinking Wedge Using Bvp4c
Nurul Umairah Ahmad & Siti Hidayah Muhad Saleh **144**

Modeling The Common Edible Oil Price Based On US Dollar Index
Aina Arinah Riduwan & Yusrina Andu **157**

A Comparative Study Of First Order Differential Equtation By Using Butcher's Fifth Order Method And Dormand Prince Method
Norlin Anas, Rahmah Shahril, Maznah Banu Mohamed Habiboo Raman, Rasidah Buang & Rosha Mohamed **164**

The Price Of Research Octane Number 97 (RON97) In Malaysia: A Univariate Time Series Approach
Nurul Syakinah Binti Mohd Zaki & Noreha Mohamed Yusof **170**

Optimising Water Supply And Demand Allocation Using Dynamic Programming: Addressing Distribution Inefficiencies
Aiman Syazwan Ahmad & Rossidah Wan Abdul Aziz **177**

Optimizing Vehicle Parking Capacity In Narrow Space: An Analysis Of The Bingle Concept
Luqman Hakim Muhamad Kamal & Najir Tokachil **184**

Numerical Solution Of Unsteady Boundary Layer Flow Over A Stretching Sheet In A Porous Medium Using BVP4C
Nurul Ain Ashiqin binti Ahmad Tarmizi & Siti Hidayah binti Muhad Saleh **191**

Selection Of Smartphone Brands By Using Fuzzy Topsis Approach
Nurin Hannani Binti Halimi, Nor Hilaliyah Binti Mohd Jamil & Siti Nazifah Binti Zainol Abidin **205**

Predicting Factors In Financial Loss Among Malaysian Scam Victims Using Machine Learning
Nur Alisa Binti Azian & Che Norhalila Binti Che Mohamed **217**

Prediction Model Of Home Loan Eligibility Status
Danisya Adlina Ruzi, Siti Aisyah Mohamed Kher, Isnewati Ab Malek & Haslinda Ab Malek **229**

An Analysis Of Factors Affecting Automobiles Resale Value In Malaysia
Syahmina Humaira Binti Hizir & Noorezatty Binti Mohd Yusop **236**

Thermal Instability Analysis Of Convection In A Horizontal Nanofluid Layer With Free-Free Boundary Condition
Norazuwin Najihah Mat Tahir, Nurfarissha Syazwina Muhamad Hairol & Noralia Qisti Mohammad Nazri **244**

The Relationship Of Excessive Online Gaming On Academic And Lifestyle Outcomes Among University Students
Nur Alya Binti Ahmad Shafri, Nur Afiqah Aliya Binti Mohd Nizam, Anis Fathiah Binti Afandi, Hafizul Izham Bin Mohamad Rizal, Muhammad Syafiq Bin Jaafar & Suriyati Ujang **257**

Thermal Instability Analysis Of Convection In A Horizontal Nanofluid Layer With Rigid-Rigid Boundary Condition
Norazuwin Najihah Mat Tahir, Norshahirah Izati Jaaffar & Nurul Nabila Rozali **267**

Predicting Gold Market Trends Using Newton's Divided Difference Method
Nurul Athirah Soleha Mohd Zin, Nurshazrin Izzati Azat, Nor Hanim Abd Rahman & Rafizah Kechil **280**

Application Of The Runge-Kutta Fourth Order (Rk4) And Runge-Kutta Fifth Order (Rk5) In Dynamic Systems: A Focus On Electrical Circuit Analysis
Nur Nayli Afiqah Abd Razak, Faranissha Shazlyn Hasmin Farouk, Nor Hanim Abd Rahman & Mohd Fahmi Zahari **287**

Optimising Courier Delivery Routes And Time Travel Using Dijkstra's Algorithm To Find The Shortest Path
Noor Hidayah Mohd Zaki, Nurin Afiqah Bakhtiar & Nurul Syahida Zanurin **296**

Solving Heat Equation Using Finite Difference Method
Nur Zeti Azreen Nor Azam, NurHayati Mohd Nazeri, Azhar Ahmad & Norshuhada Samsudin **302**

Multi-Choice Goal Programming Approach For Multi-Objective Portfolio Optimization Problems
Aisyah Safiyah Abdul Halim, Muhammad Danial Iskandar Nizam & Mazura Mokhtar **308**

Solving Poisson's Equation Using Finite Difference Method
Nur Fatin Syuhada Mohd Hasrulnizam, Nur Syazwani Mohd Azmee, Azhar Ahmad & and Sharifah Sarimah Syed Abdullah **320**

Comparative Study Of AHP And Potential Method In Supplier Selection Decision Making
Sarah Afrina Saidi, Wan Nur Haryani Wan Mohd Ariff Rashiddin & Siti Salwana Mamat **325**

Approximate Solution Of Fractional Differential Equations Using Transformation And Perturbation Method
Hana Zaidah Zulkefli & Farah Suraya Md Nasrudin **333**

Approximate Solution Of Fractional Differential Equations Using Transformation And Perturbation Method
Hana Zaidah Zulkefli & Farah Suraya Md Nasrudin **340**

Understanding Local Food Security Awareness: Insights From The Seremban Community
Nur Adriyana Shafie & Noorezatty Mohd Yusop **347**

Analyzing Food Security Of The Households: A Case Study In Malaysia Using A Logistic Regression Approach
Julia Johari & Jaida Najihah Jamidin **356**

A Decision-Making Framework For Small Restaurant Location Selection Using The Analytic Hierarchy Process
Anis Amira Asri, Nurul Atikah Amran, Nursyamimi Mahzan & Siti Sarah Januri **364**

Forecasting Maternal Mortality Rate In Malaysia
Nurul Asyiqin Sahrulnizam & Nor Akmal Md Noh **369**

Performance Of Box Jenkins And Artificial Neural Networks (ANN) Models In Air Pollution Index In Klang
Nur Aqilah Aina Abd Latif & Nor Akmal Md Noh **376**

Paddy Production Analysis In Non-Granary Areas
Nora Binti Mohd Basir & Nur Aliah Najwa Binti Ahmad Mahyudin **387**

Home Energy Management For Home Appliances Using Knapsack Model
Fatin Nur Liyana Mohd Ridzuan & Nor Aishah Md Noh **395**

Prediction Model Of Home Loan Eligibility Status
Danisya Adlina Ruzi, Siti Aisyah Mohamed Kher, Isnewati Ab Malek & Haslinda Ab Malek **409**

Radius Of Starlikeness Of A Subclass Of Generalized Bi-Univalent Functions
Nur Suhailah Zamri & Abdullah Yahya **415**

Numerical Solution Of Nanofluid Flow Over A Permeable Stretching And Shrinking Surface Using Bvp4c Solver
Nurqasrina Ahmad Azian & Siti Hidayah Muhad Saleh **432**

Default Risk Prediction Using Merton's Model And Altman's Z-Score
Anas Ahmad, Mohd Ezzaden Jasni, Mohd Shahirul Fazli Mohamad Hatta & Norliza Muhamad Yusof **445**

The Importance of Academic Writing to Undergraduate Mathematics and Statistics Students: A Case for Student Final Year Project Proceedings

Mat Salim Selamat
Universiti Teknologi MARA Negeri Sembilan, Seremban Campus, Seremban, Negeri Sembilan
matsalimselamat@uitm.edu.my

Abstract

Academic writing is often misconstrued as a secondary skill in quantitative fields such as Statistics and Mathematics. This article argues that it is actually an essential discipline for transforming technical implementation into meaningful scholarly knowledge. This article explains the benefits of a variety of academic rigors to undergraduate students, including the development of critical thinking, professional communication skills, and research integrity. Furthermore, it highlights the advantages for student supervisors (lecturers), such as enhanced assessment capabilities and strengthening institutional reputation. These arguments are contextualized in a specific initiative, namely the publication of proceedings for final year projects in the Faculty of Computer and Mathematical Sciences UiTM Negeri Sembilan. Using a framework and empirical studies, this article argues that integrating formal writing and publishing into the curriculum is a transformative practice that prepares students for global engagement and strengthens the faculty's status as a center of academic excellence.

Keywords: Academic Writing, Undergraduate Research, Final Year Project, Proceedings Publication, Higher Education

Introduction: Beyond the Final Calculation

In the Faculty of Computer and Mathematical Sciences, excellence is traditionally measured by the elegance of proof, the accuracy of models or the predictive power of algorithms. However, separately solved equations or repositories of code that remain uninterpreted have limited impact. The true value of scientific work is only evident when it is effectively communicated. Therefore, academic writing has become an important discipline that transforms technical implementation into meaningful contributions. To achieve this goal, the Faculty of Computer and Mathematical Sciences, UiTM Negeri Sembilan to publish the proceedings of the final year project. This academic writing initiative is expected to be an engine that drives the intellectual growth of students and enhances the scholarly reputation of the faculty and supervisors.

Cultivating the Scientist's Most Vital Skill: Storytelling with Data

For Bachelor of Statistics and Bachelor of Mathematics students, the core of their work is argumentation, not just computation. A student can develop sophisticated time series

predictions or graph theory algorithms, but their value remains hidden without context and narrative. Academic writing forces students to ask: What is the real-world problem? Why is the chosen method better? How should these results be interpreted, and what are its limitations?

These skills are not optional; they are globally recognized competencies. The American Statistical Association (ASA) College Report on Assessment and Instruction in Statistics Education (GAISE) clearly lists "communicating statistical results effectively" as a key learning objective [1]. Similarly, in Europe, the Bologna Process emphasizes "communication skills in a second language (often English for academic purposes)" and "the ability to communicate with peers" as core attributes of first-degree (Bachelor) graduates. A well-written proceeding paper is direct evidence of these competencies. It demonstrates a student's ability to translate complex quantitative findings to a wider audience and these are highly sought-after skills in a global, data-driven industry.

The Proceeding as a Platform for Authentic Learning and Assessment

Publishing a proceedings will elevate the final year project from a mandatory assignment to an informative scholarly experience. This process introduces students to the basics of academic life i.e. peer review. Modeled after the international conference model organised by bodies such as the Society for Industrial and Applied Mathematics (SIAM) or the International Statistical Institute (ISI), students learn to accept constructive criticism and review their work. This is where the learning process takes place [5].

A meta-analysis by [2] found that writing activities have a significant positive impact on academic achievement because they force deeper cognitive engagement. Furthermore, a study of undergraduate research experiences found that students who participate in the writing and publication process report significant improvements in their ability to analyse data and understand the research process [3].

For supervisors, written proceedings papers are an invaluable diagnostic tool. They provide a clear window into students' reasoning, revealing strengths in logic and gaps in methodology that may not be apparent in the student's final report. This allows for targeted, high-quality feedback and transforms supervision into a more effective collaborative process.

Building Professional Portfolios and Enhancing Institutional Reputation

In an increasingly competitive global marketplace, graduates from UiTM Negeri Sembilan compete with peers from around the world. Publication in faculty proceedings is a strong differentiator. It provides tangible and verifiable evidence of a student's ability to manage complex projects from concept to communication. For those pursuing graduate studies, this evidence of early research output is a significant advantage in admissions.

For the Faculty of Computer and Mathematical Sciences, the benefits are equally significant. First, publication in proceedings is a strategic asset that displays excellence. It openly demonstrates the intellectual power and practical relevance of its program to industry partners and international academic institutions.

Second, it creates a searchable archive of student work that can be used to identify

emerging research trends (e.g., the rise of AI ethics or Bayesian methods), inform curriculum development and foster internal collaboration.

Third, publication in proceedings is about building a legacy. It documents the faculty's commitment to high-quality, research-informed education, raising its profile and reputation in the national and global academic community.

Upholding Integrity and Ethical Practice

Finally, the academic writing process instills the foundation of academic integrity. By adhering to citation conventions (e.g., APA, IEEE), students learn to respect intellectual property and acknowledge the foundational work on which their own projects are built. This acquaints them with the ethical norms of their discipline, preparing them for a professional environment where transparency and credibility are non-negotiable [4].

Conclusion: From the Laboratory to the Legacy

The project to publish undergraduate proceedings is a visionary initiative that aligns with global best practices in science and mathematics education. It recognizes that the next generation of statisticians and mathematicians must be not only technical experts but also eloquent communicators. By encouraging students to undertake academic writing, the faculty has fulfilled its responsibility to shape students into well-rounded professionals, capable of producing, defending, and disseminating their conclusions.

For Supervisors and Faculty, it enhances the quality of supervision, creates a lasting legacy of student achievement, and firmly positions the faculty as a center of academic excellence

At its core, academic writing is the discipline that ensures that students' sophisticated calculations do not end up as silent numbers on the page but are heard as clear, confident, and effective voices in the ongoing global conversation of science.

References

[1] American Statistical Association (ASA). (2016). Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report. Retrieved from https://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx

[2] Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*(1), 29–58.

[3] Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. *CBE—Life Sciences Education, 6*(4), 297-306.



[4] Löfström, E. (2016). Academic integrity: The role of moral reasoning and self-efficacy in the interpersonal context of teaching and learning. In T. Bretag (Ed.), *Handbook of Academic Integrity*. Springer Singapore.

[5] Society for Industrial and Applied Mathematics (SIAM). (n.d.). *Conferences*. Retrieved from https://www.siam.org/conferences

Mortality Modelling and Forecasting: An Application of The Lee-Carter Model on The Japanese Data

Sherlinda Dinda Lim[1], Nurul Aityqah Yaacob[2, *]
[1]*Faculty of Computer and Mathematical Sciences, UiTM Cawangan Negeri Sembilan, Kampus Seremban*
[2]*Faculty of Computer and Mathematical Sciences, UiTM Cawangan Negeri Sembilan, Kampus Kuala Pilah*
[*]aityqah@uitm.edu.my

Abstract

Accurate mortality modeling and forecasting are essential for understanding population health trends, informing public health policies, and improving demographic projections. This study applies the Lee-Carter (LC) model to Japanese mortality data from 1947 to 2022 to analyze trends and improve predictive accuracy. The dataset, obtained from the Human Mortality Database, is split into training and testing sets (90:10, 80:20, and 70:30) to evaluate model performance. To enhance projection accuracy, the LC model is integrated with the autoregressive integrated moving average (ARIMA) model. The findings reveal a steady decline in mortality rates for both genders, with female mortality following a smoother trend, while male mortality exhibits greater fluctuations. Life expectancy projections indicate a continued upward trend, with female life expectancy consistently surpassing that of males throughout the forecast period (2023–2032). Performance evaluations using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Squared Error (MSE) confirm that larger training datasets improve predictive accuracy, as demonstrated by the superior performance of the 90:10 split. The LC model effectively captures historical mortality trends, though some smoothing effects were observed. The U-shaped mortality pattern —characterized by high infant mortality, low childhood mortality, and increasing mortality with age—aligns with expected demographic trends. Additionally, economic recessions have had a stronger impact on male mortality, particularly due to increased suicide rates linked to job insecurity. Overall, this study demonstrates the LC model's reliability in mortality forecasting, highlighting its potential for guiding public health policies. Future research may explore advanced machine learning techniques to further enhance predictive accuracy.

Keywords*:* ARIMA, Demographic, Lee-Carter model, Mortality, Life expectancy

1. Introduction

Mortality is a key indicator of population well-being and is studied across multiple disciplines, including medicine, demography, public health, and social sciences. Analyzing mortality patterns and trends helps assess public health conditions, evaluate healthcare interventions, and guide policy decisions [1]. Mortality rates offer insights into disease incidence, injury patterns, lifestyle habits, and healthcare accessibility, all of which shape overall population health [2].

Through mortality analysis, researchers can identify health disparities, assess healthcare systems, and develop strategies to enhance quality of life.

Mortality studies serve several crucial purposes. First, they help identify prevalent diseases contributing to significant mortality within a population, enabling medical research and efficient resource allocation in healthcare sectors [3]. Analyzing mortality data allows researchers to track disease dynamics, detect emerging health threats early, and develop preventative measures to reduce premature deaths [4]. Furthermore, mortality trends reveal the effectiveness of healthcare interventions, aiding policymakers in focusing on strategies that yield positive health outcomes [5]. Mortality research also highlights the impact of social determinants such as socioeconomic status, geographic location, and healthcare accessibility on health disparities and life expectancy [6].

One of the earliest approaches to mortality modeling was the Gompertz model, introduced by Benjamin Gompertz in 1825. While historically significant, this model struggles to accurately predict modern mortality trends, particularly among younger populations [7]. In contrast, the Lee-Carter (LC) model incorporates age and time effects, making it more adaptable for forecasting mortality trends [8]. Booth and Tickle [9] further demonstrated that the LC model produces more accurate mortality projections by effectively capturing fluctuations in mortality rates. Although the Gompertz model retains historical relevance, it lacks the predictive power and flexibility of the LC model in contemporary mortality research.

Comparative studies of mortality forecasting techniques have evaluated a range of predictive approaches, including Bayesian methods and deep learning models [9-11]. These studies focus on comparing forecasting methods to determine their accuracy across different demographic contexts [11-12] . Identifying the most reliable forecasting techniques remains a primary goal in mortality research. Mortality forecasting has become increasingly essential in modern demographic studies, providing insights into population health trends, informing social progress, and shaping policy decisions.

The LC model, introduced by Lee and Carter [13], is one of the most widely used mortality forecasting models due to its adaptability and accuracy. It decomposes the age-time mortality rate matrix using principal component analysis, representing it as a bilinear sum of age and period effects. The time component is then projected using time-series models to estimate future mortality trends. Over the years, refinements have improved the LC model's predictive accuracy. For example, Xia [14] reformulated the LC model using a Poisson regression approach to address the assumption of constant error variance. Additionally, multi-factor variations and a generalized LC model incorporating cohort effects have been developed to enhance model fit.

The LC model's popularity stems from its simplicity, interpretability, and reliability. Unlike more complex models that require extensive computations, it remains straightforward to implement while maintaining high forecasting accuracy. The model has been applied across various countries, including Indonesia [15], India [16], Brazil [17], Hungary [18], Thailand [19], and Malaysia [20]. In the United States, Lee and Carter [13] applied the model to forecast national mortality rates. These applications demonstrate its reliability in capturing mortality trends across diverse populations and periods. The LC model remains a preferred choice among demographers and actuaries due to its consistent accuracy [21].

This study analyzes mortality trends in the Japanese population by examining historical data and identifying patterns over time. It applies to the LC model alongside the autoregressive

integrated moving average (ARIMA) model to enhance forecasting accuracy by capturing both long-term structural trends and short-term fluctuations. However, the accuracy of the LC model depends significantly on data quality, as poor data selection can result in inaccurate predictions and misinformed policy decisions. To address this, the study evaluates the impact of data selection on forecasting accuracy and assesses the LC model's predictive performance using various statistical metrics. The findings contribute to the advancement of mortality modeling techniques, offering practical insights for demographers, healthcare planners, and policymakers in preparing for future demographic shifts. The structure of this paper is as follows: Section 2 outlines the data and methodology used in the study. Section 3 presents the results and analysis. Finally, Section 4 summarizes key findings and suggests directions for future research.

2. Methodology

For this research, Japan mortality data was obtained from the Human Mortality Database (https://mortality.org/). The dataset includes central mortality rates and mid-year population estimates, categorized by individual ages up to 95+ years, spanning the years 1947 to 2022. In this study, ages above 95 were grouped as 95+ to avoid irregular rates for these ages. To assess model performance, the dataset is split into training and testing sets based on periods using three different ratios: 90:10, 80:20, and 70:30, as shown in Table 1.

Table 1: Training and Testing Data Splits (90:10, 80:20, 70:30)

Ratio	Training years	Testing years
90:10	1947–2014	2015–2022
80:20	1947–2007	2008–2022
70:30	1947–1999	2000–2022

The LC model, introduced in 1992, is expressed as follows:

$$ln\big(m(x,t)\big) = a(x) + b(x)k(t) + \varepsilon(x,t), x = 0, \dots, x_m \quad (1)$$

where $m(x,t)$ is the central mortality rate, calculated by dividing the number of individuals aged x who died in year t by the exposure to death for that age group.

$$\hat{a}(x) = \frac{\sum_{t=1}^{n} ln\,(m(x,t))}{n} \quad (2)$$

$a(x)$ is the average age-specific mortality, $b(x)$ is a deviation in the mortality caused by changes in $k(t)$, $k(t)$ is the mortality index, $\varepsilon(x,t)$ is the residual at age x and time t, and $m(x,t)$ is the age-specific death rate for the x interval and the year t [22]. These variables are independent and identically distributed, with a normal distribution $N(0,\sigma^2)$ with mean 0 and variance σ^2. The parameters $b(x)$ and $k(t)$ are estimated using singular value decomposition (SVD), while $a(x)$ is the average of ln($m(x,t)$) across time. The following restrictions were placed to ensure uniqueness of the solution [13]:

$$\sum_{x_0}^{x_m} b(x) = 1 \; and \; \sum_{t_1}^{t_n} k(t) = 0. \tag{3}$$

Applying SVD in the matrix $Z(x,t)$ [23]:

$$Z(x,t) = ln\big(m(x,t)\big) - \hat{a}(x) \tag{4}$$

produces:

$$ULV' = SVD\big(Z(x,t)\big) - L(1)U(x,1)V(t,1) + \cdots + L(X)U(x,X)V(t,X) \tag{5}$$

where L is the singular value, V is the time component, and U is the age component. Estimating $a(x)$, $b(x)$, and $k(t)$ from past age-specific death rates is the first stage in using the LC model to predict mortality. Finding the average of $\ln(m(x,t))$ across time yields estimates of $\hat{a}(x)$. By estimating the first term, the estimates of $\hat{b}(x)$ and $\hat{k}(t)$ may be produced [22], where:

$$\hat{\mathrm{b}}(\mathrm{x}) = \frac{1}{\sum_{\mathrm{x}} \mathrm{u}(\mathrm{x},1)} (\mathrm{u}_1, 1, \mathrm{u}_2, 1, \ldots, \mathrm{u}_{\mathrm{x}}, 1)^{\mathrm{n}} \tag{6}$$

$$\hat{\mathrm{k}}(\mathrm{t}) = \sigma_1 \times (\mathrm{v}_1, 1, \mathrm{v}_2, 1, \ldots, \mathrm{v}_{\mathrm{t}}, 1) \tag{7}$$

After estimating all the parameters, the forecasted and actual death rates are calculated. However, a significant discrepancy often exists between them. Therefore, it is necessary to re-estimate $k(t)$ by satisfying Eq. 2 and to estimate $\hat{b}(x)$ to ensure that the forecasted and actual death rates match [24]. Thus, the parameter estimates satisfy the following condition:

$$\sum_{\mathrm{x}=\mathrm{x}_1}^{\mathrm{x}_m} \mathrm{D}(\mathrm{x},\mathrm{t}) = \sum_{\mathrm{x}=\mathrm{x}_1}^{\mathrm{x}_m} E(x,t)\big[exp\,(\hat{\mathrm{a}}(\mathrm{x}) + \hat{\mathrm{b}}(\mathrm{x})\hat{\mathrm{k}}(\mathrm{t}))\big] \tag{8}$$

where $\sum_{x=x_1}^{x_m} \mathrm{D}(x,t)$ is the total number of deaths in year t [24].

To ensure that the predicted number of deaths closely matches the observed number of deaths for a given period, the estimates were adjusted. The ARIMA model, specifically the ARIMA (0,1,0) model initially employed by Lee and Carter [13], is widely used for modeling random walk with drift and forecasting the time-varying index $k(t)$,

$$\hat{k}(t) = \hat{k}(t-1) + \theta + \varepsilon(t) \qquad \varepsilon_t \sim N(0, \sigma_k^2) \tag{9}$$

where $\varepsilon(t)$ represents normally distributed error terms with a mean of 0 and variance σ_k^2, and θ is the drift parameter [22]. To obtain the predicted values of $m(x,t)$, the projected values of the modified $k(t)$ are substituted into Eq. (1) along with the estimated $a(x)$ and $b(x)$:

$$\hat{m}(x,n+h) = \hat{m}(x,n)exp\{\hat{b}(x)(\hat{k}(n+h) - \hat{k}(n))\} \qquad h = 1,2,\ldots \quad x = 1,2,\ldots,n \tag{10}$$

where x is the age group, n is the most recent year with available data, and h is the projection horizon.

Life expectancy is calculated by integrating the projected $k(t)$ with the estimated $a(x)$ and $b(x)$ to derive forecasted age-specific mortality rates for future years. It is assumed that the force of mortality $x(t)$ remains constant within each age group. Before computing the average life expectancy, the mortality rate $m(x,t)$ is transformed into the probability of death $q_{x,t}$.

$$q_{x,t} = \frac{2\mathrm{m}(\mathrm{x},\mathrm{t})}{1 + 2\mathrm{m}(\mathrm{x},\mathrm{t})} \tag{11}$$

Life expectancy is then determined using the equation:

$$e_x(t) = \frac{1}{2} + \sum_{n-1}^{\infty} \prod_{x=0}^{n-1} P_{x,k}(t)\,, \tag{12}$$

where $P_{x,k}(t) = \prod_{i=0}^{x-1}[1 - q_{x+i}(t)]$ [25]Click or tap here to enter text..

To evaluate the accuracy of the LC model for mortality forecasts, several statistical measures are commonly used. In this study, RMSE, MAE, MAPE, and MSE were employed as key performance metrics, defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2} \tag{13}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \tag{14}$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{15}$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2 \tag{16}$$

where A_t is the actual value, F_t is the predicted value, and n is the number of observations [26]Click or tap here to enter text..

3. Results and Discussion

Figure 4.1 shows the relationship between the logarithm of the death rate and age for Japanese males and females, revealing a similar mortality pattern for both genders. The lines are color-coded using a rainbow gradient, with red representing the earliest year and violet the most

recent. Each line represents a specific year, providing a clear visualization of mortality trends over time.

Infant mortality starts high but drops sharply after birth, influenced by socioeconomic conditions and healthcare access [27]. Mortality rates remain low and stable during childhood and early adulthood, reflecting Japan's strong public health policies, disease prevention efforts, and lifestyle interventions [28-29]. From age 40 onward, mortality gradually rises due to age-related health risks and a growing prevalence of lifestyle-related diseases such as cardiovascular conditions and diabetes [28]). After age 60, the increase in mortality becomes steeper, with greater variability among individuals aged 80 and above. Chronic diseases, particularly cancer and cardiovascular illnesses, are the leading causes of death in this age group [30]. These trends highlight improved survival rates over time, demonstrating the impact of healthcare advancements, disease prevention, and lifestyle improvements on Japan's population health.

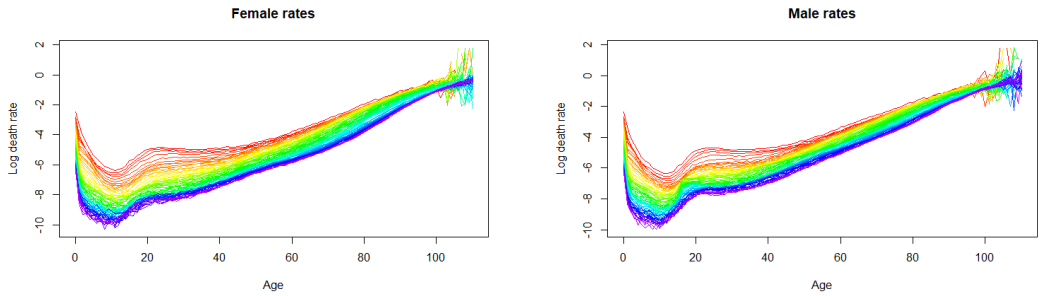


Figure 4.1: Log Death Rates Against Age

Figure 4.2 shows the relationship between log death rates and years for Japanese males and females, showing a consistent decline in mortality over time. The lines follow the same rainbow gradient, with red for the earliest year and violet for the most recent. Each line represents a specific year, providing a clear visualization of mortality trends over time. The most significant decline occurred between 1950 and 1980, slowing in later decades. Male mortality fluctuates more, particularly at older ages, while female mortality follows a smoother and more gradual decline.

Higher log death rates at the top of the graph correspond to older individuals, while lower lines indicate improvements in mortality among younger cohorts. Female mortality remains consistently lower than male mortality, aligning with the well-established trend of higher female life expectancy. Socioeconomic factors contribute to gender differences in mortality. Japan's economic stagnation from the 1990s to the early 2000s led to increased male mortality, particularly due to higher suicide rates linked to job insecurity and financial stress [31]. In contrast, women, despite facing caregiving and work-related stress, benefit from stronger social support networks, which help mitigate some of these risks [32].

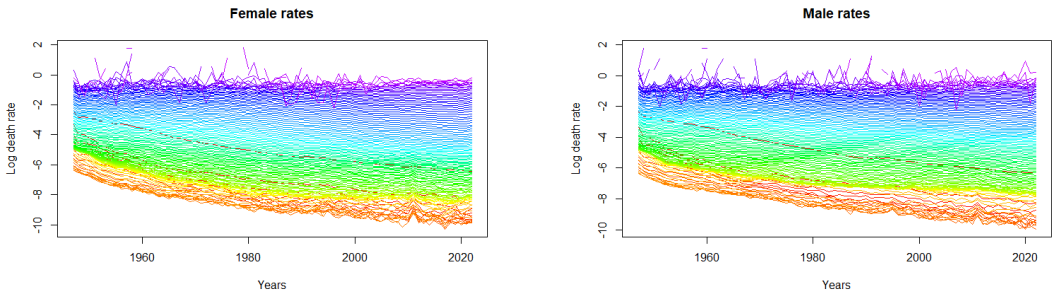


Figure 4.2: Log Death Rates Against Years

Table 4.1: Life Expectancy for the Next 10 Years

Year	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032
Female Life Expectancy	87.68	87.89	88.10	88.30	88.50	88.70	88.90	89.10	89.29	89.48
Male Life Expectancy	81.35	81.56	81.76	81.97	82.17	82.37	82.57	82.76	82.96	83.15

Table 4.1 presents a steady increase in life expectancy for Japanese males and females from 2023 to 2032. Female life expectancy remains consistently higher, reaching 89.48 years in 2032, compared to 83.15 years for males, reflecting the well-established trend of women outliving men. The 6.33-year gap suggests that lifestyle-related diseases, occupational hazards, and stress-related conditions continue to impact male mortality more significantly. While improvements in healthcare, disease prevention, and lifestyle factors contribute to the rising life expectancy for both genders, the slower increase in male longevity highlights the need for targeted health interventions to reduce preventable mortality and promote overall well-being.

Tables 4.2 and 4.3 present the performance metrics of the LC model in forecasting mortality rates for Japanese females and males using different training-testing ratios (90:10, 80:20, and 70:30). The results indicate that the model performs better with a larger training dataset, as seen in the lower RMSE, MAE, MAPE, and MSE values in the 90:10 split for both genders. For Japanese females (Table 4.2), RMSE increases from 0.5208 (90:10) to 0.9432 (70:30), indicating a rise in prediction errors as the training set decreases. Similarly, MAE rises from 0.4188 to 0.7493, MAPE from 0.0567 to 0.1041, and MSE from 0.2712 to 0.8896. This pattern suggests that reducing the training dataset leads to decreased accuracy.

A similar trend is observed for Japanese males (Table 4.3). RMSE rises from 0.2779 (90:10) to 0.4240 (70:30), while MAE increases from 0.1982 to 0.3024. Additionally, MAPE grows from 0.0302 to 0.0476, and MSE from 0.0772 to 0.1798, further confirming that a smaller training set leads to higher errors and reduced predictive reliability. These findings align with existing literature on mortality forecasting, which suggests that larger training datasets improve predictive accuracy by providing more historical patterns for learning [13]. Conversely, smaller

training data sets increase the risk of overfitting, which reduces the model's ability to generalize to new data [9].

Table 4.2: Performance Metrics for Japanese Females (90:10, 80:20, 70:30 Splits)

Performance measures	Train-test years of 90:10 ratio	Train-test years of 80:20 ratio	Train-test years of 70:30 ratio
RMSE	0.5208	0.7824	0.9432
MAE	0.4188	0.6354	0.7493
MAPE	0.0567	0.0861	0.1041
MSE	0.2712	0.6121	0.8896

Table 4.3: Performance Metrics for Japanese Males (90:10, 80:20, 70:30 Splits)

Performance measures	Train-test years of 90:10 ratio	Train-test years of 80:20 ratio	Train-test years of 70:30 ratio
RMSE	0.2779	0.3754	0.4240
MAE	0.1982	0.2683	0.3024
MAPE	0.0302	0.0400	0.0476
MSE	0.0772	0.1409	0.1798

In summary, the results indicate that the LC model performs best with larger training datasets, as shown by the higher accuracy of the 90:10 training-test split for both males and females.

4. Conclusion

This study analyzed mortality forecasting using the LC model, demonstrating its robustness and applicability to Japanese male and female mortality data. The LC model remains a benchmark in mortality forecasting, allowing comparisons with newer techniques to assess their strengths and limitations. The results highlight the role of statistical measures, such as RMSE, MAE, MAPE, and MSE, in evaluating forecast accuracy. Among these, MAE is less influenced by outliers compared to RMSE. Mortality trends in Japan show a steady rise in life expectancy, with high infant mortality in early years, stabilization in childhood, and a gradual increase in mortality with age. Female mortality consistently remains lower than male mortality, influenced by socioeconomic conditions, healthcare policies, and lifestyle factors. Economic recessions have also affected male mortality rates more significantly.

The study also examined the impact of different training-to-testing ratios (90:10, 80:20, 70:30) on model performance. The findings show that the LC model achieves higher accuracy with larger training datasets, while reducing training data leads to increased errors. The integration of the ARIMA(0,1,0) model with the LC model effectively captured historical trends and projected future mortality rates, though some smoothing effects were observed. Overall, the results align with the study's objectives, confirming an upward trend in life expectancy and a persistent gender gap in mortality, largely influenced by socioeconomic and health-related factors. The LC model remains a reliable tool for mortality forecasting, but its accuracy depends on the volume of training data. Future research could explore Bayesian methods or deep

learning techniques to improve predictive performance, especially when data availability is limited [10] and [12].

Acknowledgments

We sincerely appreciate the Human Mortality Database (HMD) (https://www.mortality.org) for providing the mortality data used in this study. We also extend our gratitude to all contributors who supported this research.

References

[1] Gebel, K., Ding, D., & Bauman, A. E. (2014). Volume and intensity of physical activity in a large population-based cohort of middle-aged and older Australians: Prospective relationships with weight gain, and physical function. *Preventive Medicine*, *60*, 131–133.

[2] Menéndez, C., Bardají, A., Sigauque, B., Sanz, S., Aponte, J. J., Mabunda, S., & Alonso, P. L. (2010). Malaria prevention with IPTp during pregnancy reduces neonatal mortality. *PloS one*, *5*(2), e9438.

[3] Amini, M., Zayeri, F., & Salehi, M. (2021). Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017. *BMC Public Health*, *21*(1).

[4] Wéber, A., Laversanne, M., Nagy, P., Kenessey, I., Soerjomataram, I., & Bray, F. (2023). Gains in life expectancy from decreasing cardiovascular disease and cancer mortality – an analysis of 28 European countries 1995–2019. *European Journal of Epidemiology*, *38*(11), 1141–1152.

[5] Raghupathi, V., & Raghupathi, W. (2023). The association between healthcare resource allocation and health status: an empirical insight with visual analytics. *Journal of Public Health (Germany)*, *31*(7), 1035–1057.

[6] Allel, K., Salustri, F., Haghparast-Bidgoli, H., & Kiadaliri, A. (2021). The contributions of public health policies and healthcare quality to gender gap and country differences in life expectancy in the UK. *Population Health Metrics*, *19*(1).

[7] Seklecka, M., Pantelous, A. A., & O'Hare, C. (2017). Mortality effects of temperature changes in the United Kingdom. *Journal of Forecasting*, *36*(7), 824–841.

[8] Booth, H., Hyndman, R. J., Tickle, L., & De Jong, P. (2006). Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research*, *15*, 289–310.

[9] Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, *3*(1–2), 3–43.

[10] Shang, H. L., Haberman, S., & Xu, R. (2022). Multi-population modelling and forecasting life-table death counts. *Insurance: Mathematics and Economics*, *106*, 239–253.

[11] Shang, H. L., & Haberman, S. (2017). Grouped multivariate and functional time series forecasting: An application to annuity pricing. *Insurance: Mathematics and Economics*, *75*, 166-179.

[12] Verbeeck, J., Faes, C., Neyens, T., Hens, N., Verbeke, G., Deboosere, P., & Molenberghs, G. (2023). A linear mixed model to estimate COVID-19-induced excess mortality. *Biometrics*, *79*(1), 417–425.

[13] Lee, Ronald D. & Carter, L. R. (1992). Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association*, *87*(419), 659.

[14] Xia, F. (2023). Why to use Poisson regression for count data analysis in consumer behavior research. *Journal of Marketing Analytics*, *11*(3), 379–384.

[15] Safitri, L., Mardiyati, S., & Rahim, H. (2019). Comparison of Lee-Carter's classic and general model for forecasting mortality rate in Indonesia. *International Journal of GEOMATE*, *16*(55), 119–124.

[16] Chavhan, R., & Shinde, R. (2016). Modeling and forecasting mortality using the Lee-Carter model for Indian population based on decade-wise data. *Sri Lankan Journal of Applied Statistics*, *17*(1).

[17] Gonzaga, M. R., Queiroz, B. L., Freire, F. H. M. A., Monteiro-da-Silva, J. H. C., Lima, E. E. C., Silva-Júnior, W. P., Diógenes, V. H. D., Flores-Ortiz, R., da Costa, L. C. C., Pinto-Junior, E. P., Ichihara, M. Y., Teixeira, C. S. S., Alves, F. J. O., Rocha, A. S., Ferreira, A. J. F., Barreto, M. L., Katikireddi, S. V., Dundas, R., & Leyland, A. H. (2024). Estimation and probabilistic projection of age- and sex-specific mortality rates across Brazilian municipalities between 2010 and 2030. *Population Health Metrics*, *22*(1).

[18] Petneházi, G., & Gáll, J. (2023). Testing the Lee-Carter model on Hungarian mortality data. *Acta Oeconomica*, *73*(1), 171–182. https://doi.org/10.1556/032.2023.00010

[19] Yasungnoen, N., & Sattayatham, P. (2016). Forecasting Thai Mortality by Using the Lee-Carter Model. *Asia-Pacific Journal of Risk and Insurance*, *10*(1), 91-105.

[20] Ibrahim, N. S. M., Lazam, N. M., & Shair, S. N. (2021). Forecasting Malaysian mortality rates using the Lee-Carter model with fitting period variants. *Journal of Physics: Conference Series*, *1988*(1).

[21] He, L., Huang, F., Shi, J., & Yang, Y. (2021). Mortality forecasting using factor models: Time-varying or time-invariant factor loadings? *Insurance: Mathematics and Economics*, *98*, 14–34.

[22] Yaacob, N. A., Pathmanathan, D., & Mohamed, I. (2022). Extending the GLM Framework of the Lee-Carter Model with Random Forest Recursive Feature Elimination Based Determinants of Mortality. *Sains Malaysiana*, *51*(7), 2237–2247.

[23] Wang, J. Z. (2007). *Fitting and Forecasting Mortality for Sweden: Applying the Lee-Carter Model*. Mathematical Statistics, Stockholm University.

[24] Azman, S., & Pathmanathan, D. (2022). The GLM framework of the Lee–Carter model: a multi-country study. *Journal of Applied Statistics*, *49*(3), 752–763.

[25] Cheng, Z., Si, W., Xu, Z., & Xiang, K. (2022). Prediction of China's Population Mortality under Limited Data. International Journal of Environmental Research and Public Health, 19(19).

[26] Rabbi, A. M. F., & Mazzuco, S. (2021). Mortality Forecasting with the Lee–Carter Method: Adjusting for Smoothing and Lifespan Disparity. *European Journal of Population*, *37*(1), 97–120.

[27] Okui, T. (2023). Association between infant mortality and parental educational level: An analysis of data from Vital Statistics and Census in Japan. *PLoS ONE*, *18*(6 June).

Petneházi, G., & Gáll, J. (2023). Testing the Lee-Carter model on Hungarian mortality data. *Acta Oeconomica*, *73*(1), 171–182.

[28] Adair, T., Kippen, R., Naghavi, M., & Lopez, A. D. (2019). The setting of the rising sun? A recent comparative history of life expectancy trends in Japan and Australia. *PLoS ONE*, *14*(3).

[29] Tanaka, H., Togawa, K., & Katanoda, K. (2023). Impact of the COVID-19 pandemic on mortality trends in Japan: A reversal in 2021? A descriptive analysis of national mortality data, 1995-2021. *BMJ Open*, *13*(8).

[30] Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., Cercek, A., Smith, R. A., & Jemal, A. (2020). Colorectal cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, *70*(3), 145–164.

[31] Okada, M., Hasegawa, T., Kato, R., & Shiroyama, T. (2020). Analysing regional unemployment rates, GDP per capita and financial support for regional suicide prevention programme on suicide mortality in Japan using governmental statistical data. *BMJ Open*, *10*(8), e037537.

[32] Nakamoto, M., Nakagawa, T., Murata, M., & Okada, M. (2021). Impacts of dual-income household rate on suicide mortalities in japan. *International Journal of Environmental Research and Public Health*, *18*(11).

AIR QUALITY ASSESSMENT IN KLANG USING PRINCIPAL COMPONENT ANALYSIS

Nur Afrina Syamimi Norazman[1], **Isnewati Ab Malek**[1,*]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, 70300 Negeri Sembilan*
[*]isnewati@uitm.edu.my

Abstract

Air pollution, a major environmental concern, negatively affects human health and the environment. In Malaysia, urban areas such as Klang in Selangor have suffered from increased air pollution due to heavy industrial activities, dense populations, and high vehicular traffic, which are particularly vulnerable. This study identified Selangor's air quality trends using the Department of Environment (DOE) data. The monitoring station in Klang was selected based on six air pollutants (sulphur dioxide (SO_2), particulate matter below 10 microns (PM_{10}), ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO) and particulate matter below 2.5 microns ($PM_{2.5}$) during three years from January 1, 2020, to December 31, 2022. This study aims to use the principal component analysis (PCA) method to classify the variables that predict the air pollution index (API) in Klang, Selangor. PCA was used to reduce dimensionality and identify significant pollutant predictors, with Bartlett's test and KMO measure supporting data suitability and rotating PCA reducing predictor variables. As a result, two principal components that highlighted PM_{10}, $PM_{2.5}$ and O_3 as the most important pollutants in Klang were revealed after the PCA was rotated using varimax rotation. The finding could assist DOE management in identifying the types of pollutants responsible for air pollution.

Keywords: Air Pollution Index, Principal Component Analysis (PCA), Klang

Introduction

The issue of air pollution is not new to any country worldwide, including Malaysia which is one of the most serious environmental issues facing the world today. There are plenty of different types of pollution, including noise, ground, water, and air pollution. Air pollution refers to the contamination of the atmosphere by another harmful substance. Air pollution is the contamination of the indoor or outdoor environment by any chemical, physical or biological agent that modifies the natural characteristics of the atmosphere [1]. After considering the information regarding air pollution, it is not only dangerous for human health but also makes it unsuitable for people to live in if air pollution continues to increase in the future.

Motor vehicles, power plants, and industry as major sources of air pollution in Malaysia [2]. Industrial smoke, cigarette smoke, open burning, and household combustion devices contribute to air pollution. Moreover, [3] said that urban areas in Malaysia have the highest levels of PM_{10}, a harmful inhalable particle, and higher levels of (O_3), a secondary air pollutant produced by industrial emissions and motor vehicle exhausts. Malaysia has also seen an increase in carbon monoxide (CO) emissions due to fossil fuel reliance.

The US Environmental Protection Agency's Air Quality measure (AQI) is the most extensively used global air quality measure, adopted by many countries, including Malaysia [4]. Based on research from [5] said that the United States Environmental Protection Agency (US EPA) provides daily air quality conditions using six criteria: sulphur dioxide (SO_2), particulate matter below 10 microns (PM_{10}), ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO) and particulate matter below 2.5 microns ($PM_{2.5}$). The index value set is classified as good, moderate, unhealthy, very unhealthy, hazardous, or emergency, indicating the air quality's potential harm to humans and the environment, as indicated in Table 1.

Table 1: Air Pollutant Index (API) of Malaysia

Air Pollution Index	Air Pollution Level
0 to 50	Good
51 to 100	Moderate
101 to 200	Unhealthy
201 to 300	Very Unhealthy
> 300	Hazardous
>500	Emergency

Seven million premature deaths annually are caused by air pollution. Malaysia's Ministry of Health reports that respiratory system diseases cause 10.36% of deaths, with 19.48% due to these diseases [1]. Pollutants contributing to air pollution include sulphur dioxide (SO_2), particulate matter below 10 microns (PM_{10}), ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO) and particulate matter below 2.5 microns ($PM_{2.5}$). Each pollutant has unique effects on human health, with sulphur dioxide SO_2 causing asthma, particulate matter below 10 microns PM_{10} causing lung cancer, ozone O_3 causing breathing problems, nitrogen dioxide NO_2 related to asthma and other respiratory diseases, and carbon monoxide (CO) causing breathing difficulties, exhaustion, dizziness, and flu-like symptoms. Particulate matter below 2.5 microns $PM_{2.5}$ can also cause cardiovascular, cerebrovascular, and respiratory effects.

Accurate prediction of API is essential for implementing timely interventions, informed decision-making, and effective air quality management. However, identifying a robust model that balances accuracy, computational efficiency, and interpretability remains a challenge. This study aims to bridge this gap by leveraging Principal Component Analysis (PCA) for dimensionality reduction. By applying these advanced techniques, the research seeks to develop an optimized model for predicting API in Klang contributing to better air quality management and public health outcomes.

Methodology

Data Description

The study used secondary data from the Department of Environment's Malaysia air pollution index, specifically the Selangor index, covering the period from January 1, 2020 to December 31, 2022. The index was calculated using daily observations from the Klang air quality monitoring stations. Pollutants included sulphur dioxide (SO_2), particulate matter below 10 microns (PM_{10}), ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO) and particulate matter below 2.5 microns ($PM_{2.5}$).

Data Processing

The data collection of the model for prediction in the current research was constructed using 1096 data received from six variables recorded daily from January 1, 2020 to December 31, 2022 from the Klang. The data were collected by the Department of Environment (DOE). Before proceeding to generate PCA, any absence of a dataset must be addressed with multiple imputations since PCA cannot handle the missing values directly. This method estimated a set of appropriate values for missing data based on the distribution of observed data.

Dimensionality Reduction

Principal component analysis (PCA) was a multivariate statistical method that combined information gathered from many different variables observed on the same individuals into a smaller number of variables known as principal components (PCs) [6]. According to [7], principal component analysis (PCA) was a dimensionality reduction (DR) technique that commonly utilised to reduce a large number of variables into a smaller number that still maintained some of the details in the large set. The PCA was performed to obtain principal components (PCs). The PCs could be expressed as

$$z_{ij} = \alpha_{i1}x1_j + \alpha_{i2}x2_j + \alpha_{i3}x3_j + \alpha_{i4}x4_j + \alpha_{i5}x5_j + \alpha_{i6}x6_j \quad (1)$$

where

z is the component score

α is the component loading
x_1 is sulphur dioxide (SO_2)
x_2 is particulate matter below 10 microns (PM_{10})
x_3 is ozone (O_3)
x_4 is nitrogen dioxide (NO_2)
x_5 is carbon monoxide (CO)
x_6 particulate matter below 2.5 microns ($PM_{2.5}$)
i is component number
j is the sample number

The correlation of variables in a PCA was evaluated using Bartlett's test of sphericity. The null hypothesis (H_0) suggested the original correlation matrix was an identity matrix, indicating unrelated variables. The alternative hypothesis (H_1) suggested the correlation matrix was significantly different from the identity matrix. A significant result (p-value < 0.05) indicated that factor analysis might be useful for the data set [8]. According to [9], this approach was used to describe the variability of a large group of connected variables and a smaller set of uncorrelated variables known as principal components. The KMO test was a measure designed to determine the suitability of data for factor analysis, measuring the adequacy of sample size. The test examined the sampling adequacy for each variable in the model and the whole model. KMO values below 0.6 indicated inadequate sampling and required corrective action [8].

Varimax rotation is a method used to reduce component complexity in PCs generated by the PCA, simplifying and improving the examination of variable structure. It involves raising high loads while decreasing small loads. PCs with eigenvalues greater than one were used to construct new variables, such as varimax factors (VFs) or factor loads, specifically in the varimax rotation process [9]. The Kaiser Criterion addresses maintain the required number of components. Varimax rotations require several variables aligned with conventional criteria, including non-observable, theoretical, and latent variables. Varimax factors (VFs) were the values used to calculate the correlation between variables. The values of varimax factors (VFs) could be categorised into three groups as shown in Table 2. The selection standard for this analysis is determined for VFs with absolute values over 0.75 [9].

Table 2: Category of varimax factor values

Varimax Factor Values	**Category**
> 0.75	Strong
$0.50 \geq VF \geq 0.75$	Moderate
$0.30 \geq VF \geq 0.49$	Weak

Result and Analysis

Data Processing

Multiple imputation was the method used to address the missing values. Since Klang data contained missing values, multiple imputations were performed.

Table 3: Result Variables

Result Variable	N of Valid Cases
Particulate matter below 10 microns (PM_{10})	1096
Particulate matter below 2.5 microns ($PM_{2.5}$)	1096
Sulphur dioxide (SO_2)	1096
Nitrogen dioxide (NO_2)	1096
Ozone (O_3)	1096
Carbon monoxide (CO)	1096

After processing the data using IBM SPSS Statistics, it was found that the data from Klang had missing values. In Klang, the variable nitrogen dioxide (NO_2) had two observations with missing values, while carbon monoxide (CO) had one observation. Since the data is numerical data, multiple imputations were conducted by replacing the missing values with the mean value of each variable. Table 3 presents the variables from the data in Klang after the missing values problem was resolved.

Dimensionality Reduction

After analyzing the data using IBM SPSS Statistics, the result revealed which pollutants were significant for determining the air pollution index for Klang. Table 4, presents the results of Kaiser-Meyer-Olkin (KMO) and Bartlett's sphericity tests for Klang.

Table 4: KMO and Bartlett's Test for Klang

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.652
Bartlett's Test of Sphericity	Approx. Chi-Square	3350.574
	df	15
	Sig.	0.000

The Kaiser-Meyer-Olkin (KMO) test for Klang data showed that the sampling adequacy was greater than 0.6, suggesting that all variables were suitable for further analysis. Factor loading calculations were conducted to evaluate the relationships between variables and extracted factors. Aside from that, Bartlett's test of sphericity displayed that the air pollution index data met the sphericity requirement, as the p-value was 0.000, which is less than 0.05. Consequently, the variables were determined to be connected rather than orthogonal. This indicated that PCA enabled the interpretation of data variability with less variables than the original number of variables.

Table 5 shows the results after applying varimax rotation. The number of principal components (PCs) to be retained was determined by examining the value of eigenvalues greater than one from the initial eigenvalues, which listed the six variables. According to [8], the retained components should account for at least 50% of the total variance, as recommended.

Table 5: Total Variance Explained for Klang

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variances	Cumulative %	Total	% of Variances	Cumulative %
1	**2.689**	44.825	44.825	2.686	44.765	44.765
2	**1.151**	19.191	64.016	1.155	19.250	64.016
3	0.987	16.456	80.471			
4	0.612	10.194	90.665			
5	0.511	8.524	99.189			
6	0.049	1.811	100.00			

As a result, Table 5 reveals that only two out of six principal components (PCs) were selected for Klang, as the eigenvalues were greater than one, accounting for 64.016% of the total variance in the data. Figure 1 shows the scree plots of the eigenvalues for the components in Klang, showing a decreasing trend as the number of components increases.

The study selected PC1 and PC2 for Klang due to their eigenvalues exceeding one, allowing for the creation of new variables known as varimax factors (VFs). VF values greater than 0.75 were chosen as a criterion for this study since it falls in the strong category, indicating moderate to significant highlights on eliminated components. Table 6 showed that three out of six air pollutants met the 0.75 VF threshold, with PM_{10}, $PM_{2.5}$ and O_3 being the most significant contributors.

Table 6: Rotated Component Matrix for Klang

	Component	
	1	2
PM_{10}	**0.933**	-0.127
$PM_{2.5}$	**0.920**	-0.136
CO	0.684	0.350
NO_2	0.632	0.456
O_3	0.299	**-0.826**
SO_2	0.112	0.330

After rotation for Klang, Component 1 accounted for approximately 44.765% of the variation in air quality data as shown in Table 5. Based on the varimax factor (VF) value categories outlined in methodology, Table 2, two pollutants with absolute VF values exceeding 0.75 have a strong VF value, which is PM_{10} (0.933) and $PM_{2.5}$ (0.920) for Klang. According to [10], particulate matter (PM), also known as dust or aerosols, is primarily caused by road traffic, construction, mining operations, wind erosion, agricultural activities, southwest monsoon wind, and burning biomass during the dry season in Malaysia.

Component 2, as shown in Table 5 contributes about 19.250% of the variability in air quality data. For this component, only one pollutant that contained absolute values of VF that exceed than 0.75: O_3 (-0.826) for Klang. According to [11], the rapid growth of industries and human activities in the ground-level atmosphere leads to increased levels of nitrogen oxides (NOx) and volatile organic compounds (VOC), precursors to ozone. Motor vehicles are a significant contributor to air pollution, public health concerns, and issues linked to the changing global climate.

Conclusion

In conclusion, the primary air contaminants in Klang may be determined using principal component analysis (PCA). The data on air pollution satisfied the sphericity assumption of Bartlett's test, with a p-value of less than 0.05, indicating that the variables associated with air pollution were correlated rather than orthogonal. By using the Kaiser-Meyer-Olkin (KMO) test, it was determined that all pollutants considered worthy of further investigation were acceptable and that the sampling adequacy was higher than 0.6. Two different PCs produced by rotating the PCA with varimax rotation showed that three out of six pollutants are the most significant pollutants in Klang, with the selected pollutants namely PM_{10}, $PM_{2.5}$ and O_3. The identification of these key pollutants is significant, as it provides a clear direction for policymakers, environmental agencies, and public health authorities to design targeted mitigation strategies. Such measures can reduce exposure to harmful pollutants, lower the prevalence of pollution-related diseases, and promote both public health and environmental sustainability in Selangor.

Acknowledgements

First and foremost, a big thank you to everyone who helped on this project. Your efforts, regardless of how small, have been important to the success of this study. A sincere thank you goes to the Universiti Teknologi MARA, Seremban Campus, Negeri Sembilan branch, for their important assistance and help which made the teamwork required for this project much simpler. The Department of the Environment (DOE) deserves credit on the tireless efforts to share accurate and comprehensive data, which significantly enhanced the accuracy of the study and results.

References

[1] World Health Organization. (2023). Monitoring air pollution levels is key to adopting and implementing global air quality guidelines. https://www.who.int/news/item/10-10-2023-monitoring-air-pollution-levels-is-key-to-adopting-and-implementing-who-s-global-air-quality-guidelines.%20Accessed%202024%20April%2027

[2] Ku Yusof, K., Azid, A., Sani, M. S. A., Samsudin, M. S., Amin, S., Rani, N., & Jamalani, M. A. (2019). The evaluation on artificial neural networks (ANN) and multiple linear regressions (*mlr*) models over particulate matter (*pm*10) variability during haze and non-haze episodes: A decade case study.

[3] Astro Awani (2024). Vehicle emissions are polluting malaysia's cities. https://www.astroawani.com/berita-malaysia/vehicle-emissions-are-polluting- malaysias-cities-453313

[4] Mohamed, N., Sulaiman, L. H., Zakaria, T. A., Kamarudin, A. S., & Rahim, D. A. (2016). Health risk assessment of *pm*10 exposure among school children and the proposed API level for closing the school during haze in Malaysia. *International Journal of Public Health Research*, *6*(1), 685–694.

[5] Maleki, H., Sorooshian, A., Goudarzi, G., Baboli, Z., Tahmasebi Birgani, Y., & Rahmati, M. (2019). Air pollution prediction by using an artificial neural network model. *Clean Technologies and Environmental Policy*, *21*, 1341–1352.

[6] Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*(1), 100.

[7] Hasan, B. M. S., & Abdulazeez, A. M. (2021). A review of principal component analysis algorithm for dimensionality reduction. Journal of Soft Computing and Data Mining, 2(1), 20–30.

[8] Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, 9(1), 4–11.

[9] Mahmud, N., Zulkifli, N. E. S., & Muhammad Pazi, N. S. (2021). A study on air pollution index in Sabah and Sarawak using principal component analysis and artificial neural network. *Voice of Academia (VOA)*, 17(1), 20–29.

[10] Sentian, J., Jemain, M. A., Gabda, D., Franky, H., & Wui, J. C. H. (2018). Long-term trends and potential associated sources of particulate matter (PM_{10}) pollution in Malaysia. *WIT Transactions on Ecology and the Environment*, *230*, 607–618.

[11] Hashim, N. I. M., Yusoff, N. A. I. M., Noor, N. M., & Ul-Saufie, A. Z. (2019). Assessment of surface ozone concentration in northern peninsular malaysia. *IOP Conference Series: Materials Science and Engineering*, 551(1), 012100.

Analyzing Attrition Among the Married Employees

Amanda Balqish Alri Mohammad Haelmi Alri[1], Noreha Mohamed Yusof[1,*]
[1]*Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, 70300, Seremban, Negeri Sembilan, Malaysia.*
[*]noreh144@uitm.edu.my

Abstract

Employee attrition, a major challenge for organizations globally, is the phenomenon of employees leaving the company for a variety of reasons. This paper explores the scenario of employee attrition among married employees using descriptive statistical analysis. Using a dataset of 366 married employees from the Kaggle website, the study applies descriptive statistical analysis to evaluate the significance of variables such as age, gender, total working years, monthly income, distance from home, work-life balance, and job satisfaction. Descriptive statistics revealed that the average number of participants is 39.35 years with total working years ranging widely between 1 and 50 years. Monthly income also varies significantly, with an average of $7,410.37, reflecting a diverse sample. Employees who live farther from work exhibit higher attrition rates, with an average commute distance of 50.36 km. Furthermore, gender analysis highlights a greater propensity for females to leave their roles compared to males. These findings highlight the necessity for organizations to implement targeted retention strategies, focusing on improving work-life balance and addressing gender disparities. This study contributes to the field of employee retention by offering actionable insights to guide organizational decision-making, thus enhancing workforce stability and performance.

Keywords*:* employee attrition, married employees, descriptive analysis, work-life balance, job satisfaction, gender disparities

Introduction

In today's business world, employee attrition causes substantial issues and long-term problems for organizations, reducing their competitiveness [1]. According to Khan [2], attrition describes the process of gradually decreasing the number of employees due to several reasons such as voluntary resignation, retirement, or termination. Meanwhile, [3] stated that employee attrition is the people who progressively leave the organization for a variety of reasons. Employee attrition is typically categorized into four types which are voluntary attrition, involuntary attrition, compulsory attrition, and natural attrition [4]. These categories help organizations identify the underlying causes of employee attrition and design effective solutions to reduce attrition and boost employee retention.

According to [5], voluntary attrition refers to employees' leaving an organization out of their own will, which is motivated by factors such as dissatisfaction with one's job, lack of opportunities, or better job options. For instance, [6] using data from United States, found that those employees who are very satisfied with their jobs and have a good work-life balance tend to have lower intentions of voluntary attrition. Similarly, a study by [7] pointed out that

employees who feel unappreciated or poorly compensated have a tendency to quit the job more willingly.

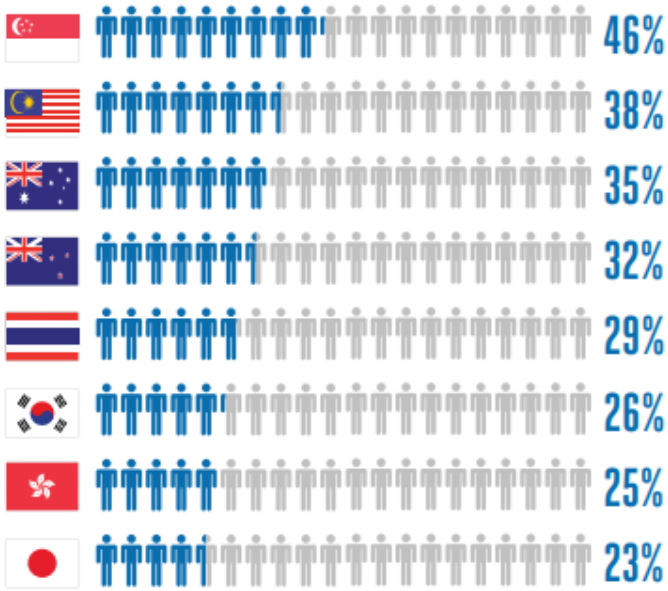


Figure 1: Employees Who Intend to Leave Their Current Job within 12 Months.

Figure 1 by [8] illustrates the percentage of employees who intend to leave their current position within the next 12 months in eight countries: Singapore, Malaysia, Australia, New Zealand, Thailand, South Korea, Hong Kong and Japan. Based on Figure 1, Malaysian workers are the second in the Asia-Pacific region. Pull factors include better pay, job opportunities, and a good work-life balance.

This paper explores the trend of the employee attrition among married employees using descriptive analysis. The variables considered are age, gender, total working years, monthly income, distance from home, work-life balance, and job satisfaction.

Methodology

The dataset "Employee Attrition and Factors" from the Kaggle website [9] was used. The variables selected for descriptive analysis include age, gender, total working years, monthly income, distance from home, work-life balance and job satisfaction. This study stands out from other existing research on employee attrition due to its exclusive focus on married employees. This study only analysed attrition among married employees, hence examining a randomly selected sample of 366 from the total 7511 cases of married employees.

Result and Discussion

This subtopic presents the findings of this investigation based on the data collected for this study. The analysis includes findings from descriptive statistics analysis. All statistical analyzes were performed with IBM SPSS version 26.

- **Age**
  The age of employees is a significant variable to consider when analyzing employee attrition patterns. Table 1 presents a summary of descriptive statistics for the age variable of the study participants.

Table 1: Summary of Age Variable of Employee Attrition

Variable	Minimum	Maximum	Mean	Standard Deviation
Age	18	59	39.35	12.25

Based on Table 1, the mean age is 39.35 years, and the standard deviation is 12.25. The individuals' ages range from 18 to 59. This indicates that married employees in this study are typically in their late 30s, which corresponds to a mid-career stage for most individuals. This age range implies that the workforce consists of both relatively young employees (early in their careers) and elderly employees who are possibly approaching retirement.

- **Gender**
  Gender plays an important role in evaluating employee turnover patterns since it can highlight potential inequalities in workplace experiences and retention. The following graph represents attrition rates by gender.

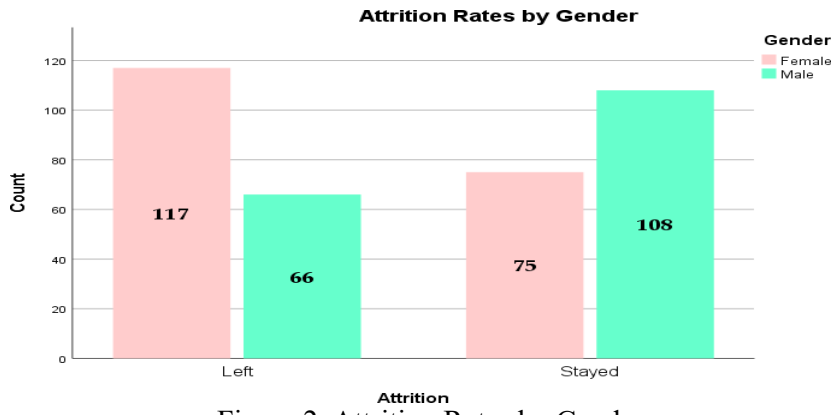


Figure 2: Attrition Rates by Gender

Based on Figure 2, the bar chart illustrates the attrition rates of married employees classified by gender (male and female). It reveals that among those who left their jobs, 117 were females and 66 were males. In contrast, among those who stayed, there were 75 females and 108 males. This indicates that a higher proportion of females among married employees left their positions compared to males, while males showed a greater tendency to stay in their roles. The data highlights a gender imbalance in employment retention among married individuals.

- **Total Working Years**
  Table 2 summarizes the descriptive statistics for the total working years among the study participants. This variable provides insights into employee tenure and its potential impact on attrition trends.

Table 2: Summary of Total Working Years of Employee Attrition

Variable	Minimum	Maximum	Mean	Standard Deviation
Total Working Year	1	50	15.80	11.95

Table 2 shows the individuals' tenure at the organization range greatly, with an average of 15.80 years and a standard deviation of 11.95, ranging from 1 year to 50 years. This shows a range of recent hires and seasoned workers, indicating a significant degree of tenure variety in the workforce. A longer average tenure among married employees may indicate loyalty, but this group may also be more likely to leave if they reach retirement age or seek new possibilities after a long period of service.

- **Monthly Income**
  Monthly income is an important factor in understanding employee attrition because it directly affects job satisfaction and retention. Table 3 summarizes the descriptive data for the monthly income variable among the study participants.

Table 3: Summary of Monthly Income of Employee Attrition

Variable	Minimum	Maximum	Mean	Standard Deviation
Monthly Income	1226	13840	7410.37	2235.43

Based on Table 3, monthly income has an average of $7,410.37 and a standard deviation of $2,235.43, ranging from $1,226 to $13,840. This large variation most likely reflects variations in job functions, levels, and responsibilities within the company.

- **Distance from Home**
  Table 4 summarizes the descriptive statistics for the distance from home. This statistic captures employees' commuting distance, which is a crucial aspect in determining potential implications for employee attrition.

Table 4: Summary of Distance from Home of Employee Attrition

Variable	Minimum	Maximum	Mean	Standard Deviation
Monthly Income	1	99	50.36	29.27

According to Table 4, the distance from home for employee attrition ranges from 1 to 99 km. The average distance is 50.36 km, with a standard deviation of 29.27 km. This shows that employees live an average of 50 km away from their employment, with significant variation between people. The large range and high standard deviation suggest that some employees live quite close to the business, while others travel from

much further away. This disparity in commute distances could influence employee attrition.

- **Work-Life Balance**
  Work-life balance is an important component in determining employee happiness and retention. The stacked bar graph depicts the distribution of attrition rates.

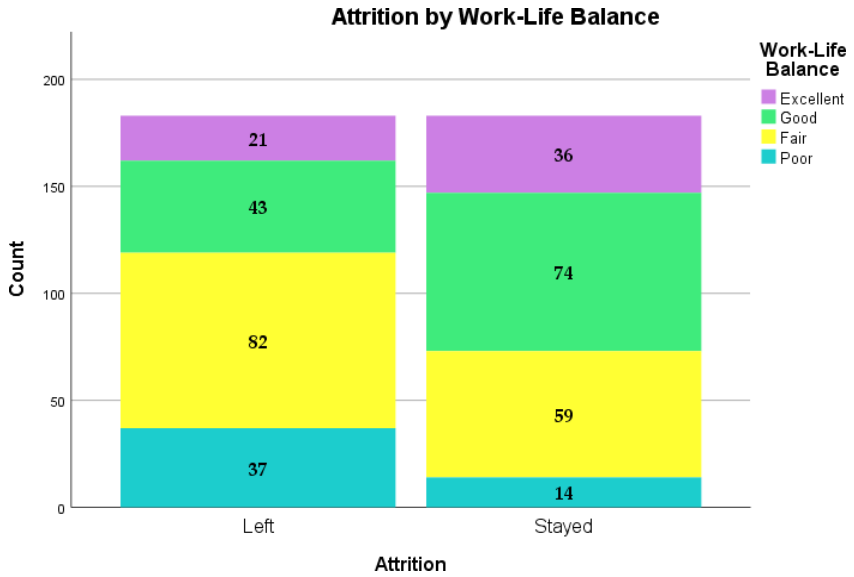


Figure 3: Attrition by Work-Life Balance

Figure 3 shows that the majority of employees who departed reported a 'fair' work-life balance (82), followed by 'good' (43), 'poor' (37), and only a small minority with 'excellent' (21). In contrast, among those who stayed, the majority reported having a 'good' work-life balance (74), followed by 'fair' (59), 'excellent' (36), and very few with 'Poor' (14). This implies that employees with a higher work-life balance (good or excellent) are more likely to stay, whereas those with a worse work-life balance (fair or poor) are more likely to depart. These findings emphasize the necessity of promoting a work-life balance in order to increase employee retention.

- **Job Satisfaction**
  The following graph shows how job satisfaction levels (low, medium, high, and very high) categorize employee attrition. It illustrates the distribution of employees who stayed or left the organization based on various satisfaction categories.

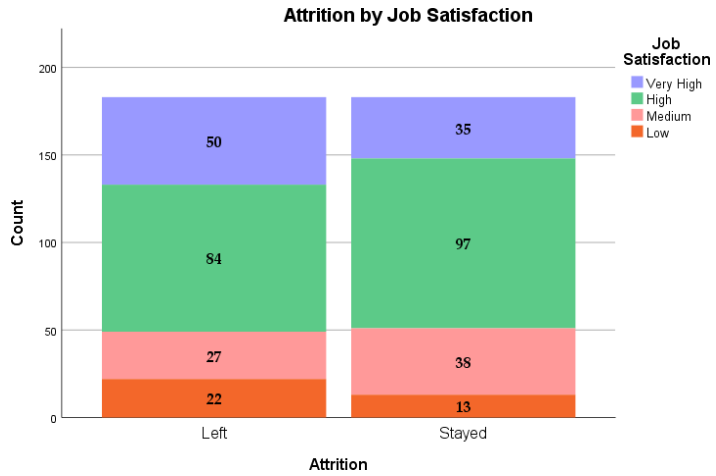


Figure 4: Attrition by Job Satisfaction

Figure 4 shows that among those that left, the majority reported medium work satisfaction (84), followed by high satisfaction (50). Similarly, the majority of employees who stayed indicated moderate satisfaction (97), followed by high satisfaction (35). Interestingly, a significant proportion of employees with extremely high satisfaction left (27), whereas fewer with poor satisfaction remained (13). This implies that, while medium contentment is the most common, attrition is not limited to lower satisfaction levels, underlining the possible impact of other factors.

Conclusion and Recommendations

Descriptive statistics analysis provides some important determinants of employee attrition in the case of married employees. The analysis shows that age, gender, total working years, monthly income, distance from home, work-life balance, and job satisfaction are significant towards attrition. Precisely, the probability of quitting is higher for younger employees, females, employees with poor work-life balance, and employees with medium job satisfaction. Another critical factor is the distance from home, where a longer commute shows higher attrition. These insights put into perspective that organizations need to address these factors to reduce employee turnover and improve retention.

To reduce attrition and promote a stable workforce, organizations should implement flexible work arrangements to enhance work-life balance, as employees with poor balance are more like to leave. Addressing gender disparities by introducing mentorship programs, gender-inclusive policies, and leadership opportunities for women can also help retain talent. Regular employee feedback surveys, career development programs, clear promotion pathways, and recognition efforts can improve job satisfaction. For employees with long commutes, offering transportation allowances, relocation assistance, or closer work options can help minimize attrition. Additionally, focusing on younger employees by providing growth opportunities, training, and

competitive compensation packages can help retain them. These strategies target key drivers of attrition and contribute to a more engaged workforce.

References

[1] F. Alsubaie and M. Aldoukhi, "Using machine learning algorithms with improved accuracy to analyze and predict employee attrition," *Decision Science Letters*, vol. 13, no. 1, pp. 1–18, 2024.

[2] S. Y. Khan, "Study on the most determining factor of employee attrition: Age factor," *International Journal of Engineering Research & Technology (IJERT) NCRIETS*, vol. 7, no. 12, 2019.

[3] S. Ganapathisamy and V. Narayan, "A long short-term memory with recurrent neural network and Brownian motion butterfly optimization for employee attrition prediction," *International Journal of Intelligent Engineering & Systems*, vol. 17, no. 1, 2024.

[4] G. Negi, "Employee attrition: Inevitable yet manageable," *International Monthly*, 2013.

[5] H. Alqahtani, H. Almagrabi, and A. Alharbi, "Employee attrition prediction using machine learning models: A review paper," *International Journal of Artificial Intelligence & Applications*, vol. 15, pp. 23–49, 2024.

[6] A. Qutub, A. Al-Mehmadi, M. Al-Hssan, R. Aljohani, and H. S. Alghamdi, "Prediction of employee attrition using machine learning and ensemble methods," *International Journal of Machine Learning and Computing*, vol. 11, no. 2, pp. 110–114, 2021.

[7] İ. E. Kaya and O. Korkmaz, "Machine learning approach for predicting employee attrition and factors leading to attrition," *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi*, vol. 36, no. 4, pp. 913–928, 2021.

[8] Workday, "IDC infographic – The drive for the digitally-enabled workforce," [Online]. Available: https://www.workday.com (n.d.).

[9]Kaggle, "Find open datasets and machine learning projects," [Online]. Available: https://www.kaggle.com/datasets. [Accessed: May. 25, 2024].

ANALYSIS OF WAITING LINES AND SERVICE FACILITIES: A STUDY USING M/M/1 AND M/D/1 QUEUING MODELS

Stephanie Morissa Anak Lister Neddi[1] and Najir Tokachil[2,*]

[1,2]College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan.
*najir2992@uitm.edu.my

Abstract

The increase in vehicles nowadays has led to many problems on the road, especially traffic congestion. Traffic congestion usually occurs during the busy hour, for example, in the morning and evening. This might happen because of the busy city and fewer people want to commute using public transport. Therefore, this project aims to analyse the waiting lines and service facilities by applying two queuing models, M/M/1 and M/D/1, to overcome traffic congestion. The project took place at the traffic light located in front of UiTM Seremban Campus, where the data is collected during the morning time, which is from 7.00 am to 8.00 am. The data collection is conducted for four types of traffic, assigned as Traffic A, B, C, and D. The project's primary focus is to analyse the performance of M/M/1 and M/D/1 models regarding waiting times and queue lengths. This project will compare the performance of both models under similar conditions, such as the duration of traffic lights. As a result, this project has proven that the M/D/1 model consistently outperformed another model, M/M/1, in reducing traffic delays. M/D/1 has displayed a lower average for vehicle count and waiting time in the system and queue. In the future, the consideration of vehicle type can be highlighted because it could directly influence the waiting time at traffic lights.

Keywords: Queuing models, M/M/1 model, M//D/1 model

Introduction

Traffic congestion has been frequently observed to be the major problem related to mobility in every country, and it has led to many problems for everyone, such as waste of time and energy and fuel consumption [1]. Queuing theory might help overcome this problem. The queuing theory is used to analyze the queues, where Danish Engineer A.K. Erlang founded the queueing theory in 1913 [2]. Queuing is generally used to address a few issues, typically related to balancing economics and effectiveness, requiring waiting and delays during serving customers or the maintenance of equipment [3]. In 2022, Shanmugasundaram and Sivaram mentioned that in a few queuing models, some preparations needed to be done before initiating a service, where this preparation is done for customers that might happen in hospitals, processes of production, banks, etc. [4].

However, a roadway's traffic characteristics are influenced by various factors, such as the surface type, roadway width, driver skill, side activities, road maintenance, etc. [5]. Thus, this project uses M/M/ and M/D/1 queuing models in operations research and performance analysis to describe systems within a traffic light. The M/M/1 model consists of a single server, where arrivals follow a Poisson process (Markovian, or "M"), service times are exponentially distributed (Markovian), and there is only one server (1). The queue operates under a first-

come, first-served (FCFS) discipline, and there is no limit on queue length. The M/D/1 model extends this by having D parallel servers instead of just one, allowing multiple customers to be served simultaneously. The presence of multiple servers reduces waiting times and congestion compared to the M/M/1 system, making it more efficient for handling higher traffic loads.

Therefore, to go into detail about this model, the traffic light in front of UiTM Kampus Seremban was selected for this project. The data collection was done during peak hours in the morning. It involved four types of traffic: Traffic A, B, C, and D. This project aimed to demonstrate the average vehicle count and time of waiting in the system and the queue by utilising both models. Another purpose of this analysis is to compare the two models by considering waiting times and the quantity of vehicles passing through the traffic signal.

Methodology

This section indicates the method implemented in this project, including data collection and model formulation. It also explains the illustration of data collection in detail.

Illustration of The Traffic Light System Project at Persiaran Seremban Tiga

Figure 1.1 shows the traffic light system project at Persiaran Seremban Tiga. The traffic lights in front of the Universiti Teknologi MARA (UiTM) Campus Seremban have provided four traffic lights. The first traffic assigned is Traffic A, which meets towards Rasah Jaya. The second traffic, known as Traffic B, meets towards the entrance gate of UiTM Seremban for vehicles that arrive from the University Commercial Centre. The data for Traffic A and Traffic B were taken over a 100-meter distance. Next, the third traffic was assigned as Traffic C, measured for vehicles arriving from Rasah Jaya towards Unicity Mall. Finally, the fourth traffic is measured for vehicles arriving from Rasah Jaya and approaching the UiTM Seremban entrance gate. As for Traffic C and Traffic D, the data was taken for 50 meters. Therefore, in this project, the data was taken during busy hours from 7.30 a.m. to 8.00 a.m., and there were about 5 times of traffic lights that were taken for 30 minutes. The arrival rate of vehicles for Traffic A, B, C and D is assigned as $\lambda_A, \lambda_B, \lambda_C, \lambda_D$ Respectively. The service rate of vehicles for Traffic A, B, C, and D is assigned as $\mu_A, \mu_B, \mu_C, \mu_D$ respectively.

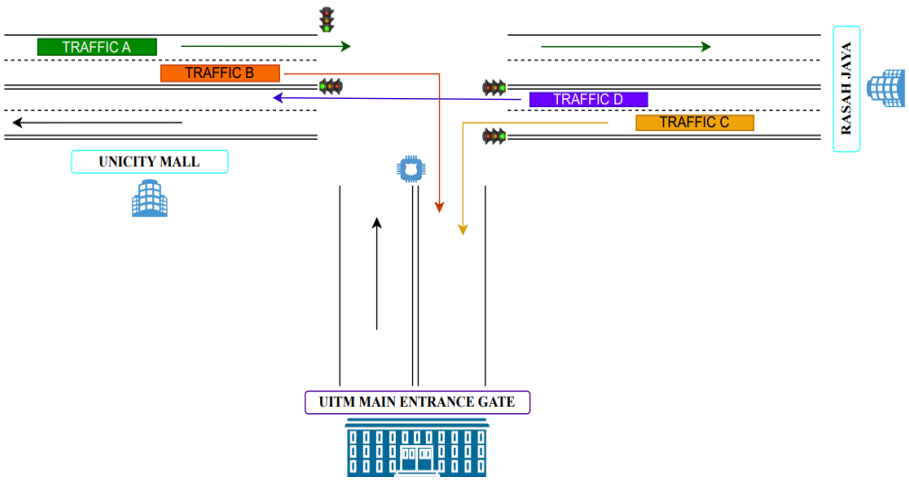


Figure 1.1 Illustration of Traffic Light in Front of UiTM Seremban Campus

The Data collection

Table 1.1 displays the data obtained during the morning session for an hour in Traffic A.

Table1.1 Data Collection of Traffic A

Traffic A	Arrival number of vehicles, λ_A	Service number of vehicles, μ_A	Waiting time (minutes)	Service time (minutes)
1	3	7	0.17	0.29
2	1	4	0.17	0.47
3	2	7	0.17	0.52
4	1	3	0.17	0.36
5	1	7	0.17	0.50

Table 1.2 displays the data observed during the one-hour morning session at Traffic B.

Table 1.2 Data Collection of Traffic B

Traffic B	Arrival number of vehicles, λ_B	Service number of vehicles, μ_B	Waiting time (minutes)	Service time (minutes)
1	12	15	0.49	0.42
2	15	18	0.49	0.44
3	10	17	0.49	0.42
4	15	19	0.49	0.51
5	13	20	0.49	0.55

Table 1.3 below displays the data from an hour of observation in the morning session at Traffic C.

Table 1.3 Data Collection of Traffic C

Traffic C	Arrival number of vehicles, λ_C	Service number of vehicles, μ_C	Waiting time (minutes)	Service time (minutes)
1	4	12	0.50	0.44
2	5	8	0.50	0.30
3	4	7	0.50	0.44
4	3	7	0.50	0.44
5	3	6	0.50	0.23

Table 1.4 exhibits the data obtained during the morning session for an hour at Traffic D.

Table 1.4 Data Collection of Traffic D

Traffic D	Arrival number of vehicles, λ_D	Service number of vehicles, μ_D	Waiting time (minutes)	Service time (minutes)
1	12	16	1	0.45
2	7	9	1	0.28
3	11	11	1	0.40
4	9	11	1	0.43
5	9	9	1	0.41

Thus, Tables 1.1 to 1.4 show that the data collected are applied using the two queuing models, M/M/1 and M/D/1, to determine the waiting times and queue lengths.

The Formulation of the Mathematical Model

Table 2.1 below summarises the formulas for the M/M/1 and M/D/1 models.

Table 2.1 Formulas for M/M/1 and M/D/1 model

Terms	M/M/1	M/D/1
Average output rate, ρ	$\rho = \frac{\lambda}{\mu}$	$\rho = \frac{\lambda}{\mu}$
Average Vehicles Count in the System, L_s	$L_s = \frac{\lambda}{\mu - \lambda}$	$L_s = \frac{2\lambda\mu - \lambda^2}{2\mu(\mu - \lambda)}$
Average Vehicles Count in the Queue, L_q	$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$	$L_q = \frac{\lambda^2}{2\mu(\mu - \lambda)}$
Average Time of Waiting in the System, W_s	$W_s = \frac{1}{\mu - \lambda}$	$W_s = \frac{2\mu - \lambda}{2\mu(\mu - \lambda)}$
Average Time of Waiting in the Queue, W_q	$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$	$W_q = \frac{\lambda}{2\mu(\mu - \lambda)}$

Results and Discussions

In this section, the table below provides the average vehicle count in the system, L_s Traffic A, B, C, and D were calculated using the M/M/1 and M/D/1 queuing models.

Average Vehicle Count in Queue, L_q

The average vehicle count in the system for both queuing models for all four traffic lights has been compared, as displayed in Table 3.1.

Table 3.1 Average Vehicle Count in the System, L_s

Traffic	**M/M/1**	**M/D/1**
Traffic A	2.81	3.04
Traffic B	2.70	1.22
Traffic C	0.59	0.58
Traffic D	0.54	0.47

According to Table 3.1, Traffic A had the most significant number of vehicles on average, 2.81 in M/M/1, whereas 3.04 in the M/D/1. Traffic B likewise shows a relatively high value of 2.70 in M/M/1 for the vehicle count on average but a lower value of 1.22 in M/D/1. Following that, there are fewer vehicles in Traffics C and D, with Traffic C recording 0.59 and 0.58 for M/M/1 and M/D/1, respectively. The lowest average, 0.54 in M/M/1 and 0.47 in M/D/1 was obtained by Traffic D. These results indicate how the M/D/1 queuing model can reduce the vehicle quantity in the system by handling traffic more effectively than the M/M/1 model.

Average Vehicle Count in Queue, L_q

Table 3.2 provides the average vehicle count in a queue, L_q for Traffic A, B, C, and D.

Table 3.2 Average Vehicle Count in Queue, L_q

Traffic	M/M/1	M/D/1
Traffic A	2.15	1.84
Traffic B	2.00	0.59
Traffic C	0.23	0.16
Traffic D	0.20	0.11

Based on Table 3.2, Traffic A records an average of 2.15 vehicles in the M/M/1 model, whereas it records 1.84 vehicles in the M/D/1 model. The smaller number of cars indicated in M/D/1 compared to M/M/1 shows that M/D/1 has a constant service rate. Nevertheless, there is a noticeable difference between the two models in Traffic B. After applying the M/M/1 model, it acquired an average of 2.00 vehicles; however, only 0.59 vehicles were in the M/D/1 model. Thus, the M/D/1 model has presented a lower average vehicle count, which might cause this model to have a deterministic rate of service that shortens the queue length for Traffic B. Compared to Traffic A and B, there are typically fewer cars queued up in Traffic C and D. Following Traffic C, the vehicle count average of 0.16 in M/D/1 and 0.23 in M/M/1 was obtained. However, in Traffic D, it obtained the value for vehicle number average of 0.20 and 0.11 in M/M/1 and M/D/1, respectively. After observing all the results, M/D/1 still indicates that it has obtained a lower average vehicle count in the queue than the M/M/1 model. Consequently, it is proven that the M/D/1 model consistently displays, in the queue, that it has fewer vehicles waiting than the M/M/1 model across all four traffic types. In conclusion, Traffic B for M/D/1 has a significantly lower number of vehicles in the queue.

Average Time of Waiting in System, W_s

Table 3.3 shows the average time of waiting in the system, W_s for Traffic A, B, C and D.

Table 3.3 Average Time of Waiting in System, W_s

Traffic	M/M/1	M/D/1
Traffic A	0.32	0.25
Traffic B	0.10	0.05
Traffic C	0.08	0.08
Traffic D	0.06	0.05

Table 3.3 shows that Traffic A experienced a longer waiting period than M/M/1, with an average waiting time of 0.32 seconds. However, Traffic A obtained a slightly shorter waiting time in M/D/1, 0.25 seconds. For Traffic B, the average difference is quite significant compared to Traffic A, where in M/D/1, it indicates a much shorter waiting time of 0.05 seconds; however, the M/M/1 model only obtained an average of 0.10 seconds. For Traffic C, both M/M/1 and M/D/1 models obtained the same waiting time of 0.08 seconds, respectively. However, for Traffic D, there was not much difference in average waiting time between the two models. Therefore, due to its deterministic character, the M/D/1 model has consistently shown that it can also reduce the waiting time in all traffic.

Average Time of Waiting in Queue, W_q

Table 3.4 shows the queue's time of waiting, W_q for Traffic A, B, C, and D, which were calculated using a two-queuing model named M/M/1 and M/D/1.

Table 3.4 Average Time of Waiting in Queue, W_q

Traffic	M/M/1	M/D/1
Traffic A	0.24	0.17
Traffic B	0.07	0.02
Traffic C	0.03	0.02
Traffic D	0.02	0.01

Table 3.4 indicates that the M/D/1 model continues to give consistently reduced average queue waiting times compared to the M/M/1 model for all types of traffic. The difference between the models indicates that the M/D/1 model features a deterministic service time, which is more effective and predictable than the random character of M/M/1 service times. Higher arrival rates in Traffic A resulted in the longest waiting times, with 0.24 seconds in M/M/1 and 0.17 seconds in M/D/1. Traffic D had the lowest waiting times, at 0.02 seconds in M/M/1 and 0.01 seconds in M/D/1. This could be due to improved service effectiveness and decreased traffic intensity. On the other hand, Traffic B, C, and D gradually display lower waiting times. For Traffic D, the slight difference in waiting times between the two models shows how the impact of the model selection decreases with lower traffic intensities.

Comparison of M/M/1 and M/D1

Table 3.5 summarizes M/M/1 and M/D/1 to determine the best model.

Table 3.5 Comparison of Total Average of Parameters for M/M/1 and M/D/1 Models

Categories	M/M/1	M/D/1
Total Average Vehicles Count in the System, L_s	1.66	1.33
Total Average Vehicles Count in the Queue, L_q	1.15	0.68
Total Average Time of Waiting in the System, W_s	0.14	0.11
Total Average Time of Waiting in the Queue, W_q	0.09	0.06

Table 3.5 shows that in the M/M/1 model, the total average vehicle count in the system decreases from 1.66 to 1.33 in the M/D/1 model. These changes have demonstrated that M/D/1, which has a feature of constant time of service, achieved a better traffic flow compared to the M/M/1 model. In the queue, the total average vehicle count in M/D/1 also displayed a lower average of 0.68, while 1.15 was on the M/M/1, which is still higher than the M/D/1. The lower average vehicle count in queues for M/D/1 has highlighted the consequences of constant service rates in reducing congestion, which has successfully decreased the vehicle count lining up in traffic. Applying the M/D/1 model also obtained a lower average of 0.11 seconds, whereas the M/M/1 model obtained a higher average of 0.14 seconds; thus, there is little difference. However, the M/D/1 model still shows a lower average value up to this point. This indicates that the constant time of service at random arrival rates, which the M/D/1 model has improved efficiency in handling vehicles within the system. Next, in the queue, as for the total average time of waiting, W_q also have a lesser value in M/D/1, which is 0.06 seconds; meanwhile, in

the M/M/1 model, it obtained an average of 0.09 seconds, slightly higher than the value for the waiting time that M/D/1 has acquired. The results obtained from these values have indicated that the model of M/D/1 consistently shows a capability to reduce delays and ensure smoother traffic flow.

As a summary, the queueing model of M/D/1 has consistently outperformed the M/M/1 model in solving traffic congestion and in all the metrics evaluated. M/D/1 queueing is characterized by a deterministic service time that has successfully decreased the average vehicle count in the system and queue, along with the average time of waiting in the system and queue. Therefore, it might help reduce the congestion at the traffic light in front of UiTM Kampus Seremban.

Conclusion

This research analyses traffic congestion at the traffic light in front of UiTM Kampus Seremban using two queuing models, M/M/1 and M/D/1. Data was collected during peak morning hours across four traffic directions (A, B, C, and D) to evaluate vehicle count and waiting time. Results show that the M/D/1 model consistently outperformed M/M/1, as its constant service time led to lower waiting times and improved traffic flow compared to the random service times of M/M/1. In the future, it recommends implementing traffic management strategies based on M/D/1, such as consideration of vehicle type, which can be highlighted because it could directly influence the waiting time at traffic lights. Besides, promoting public transportation and investing in dedicated student transport can also alleviate traffic.

Acknowledgement

The authors express gratitude to all reviewers and anonymous referees for their valuable comments and recommendations, which enhanced the presentation of this article.

References

[1] Bakkal, S. E., Lakhouili, A., & Essoufi, E. H. (2022). Modeling and Simulation of Adaptive Traffic Control System for Multi-Intersection Management using Cellular Automaton and Queuing System. International Journal of Advanced Computer Science And Applications, 13(5). https://doi.org/10.14569/ijacsa.2022.01305106

[2] Kalwar, M. A., Khaskheli, S. A., Bux Marri, H., & Khan, M. A. (2021). Applications of Queuing Theory and Discrete Event Simulation in Health Care Units of Pakistan. *International Journal of Science and Engineering Investigations*, *10*. www.IJSEI.com

[3] Ekeocha, R. J. O., & Ihebom, V. I. (2018). The use of queuing theory in the management of traffic intensity. *International Journal of Sciences*, *4*(03), 56–63. https://doi.org/10.18483/ijsci.1583

[4] Shanmugasundaram, S., & Sivaram, G. (2022). A Study on Single Server Feedback Priority Queue When Catastrophes Occur. In *Advances and Applications in Mathematical Sciences* (Vol. 21, Issue 5).

[5] Prasad, K. L. M., & Usha, B. N. (2015). *A comparison between M/M/1 and M/D/1 queuing models to vehicular traffic at Kanyakumari district*. https://api.semanticscholar.org/CorpusID:96458218

Classification Symptoms of Mental Health Problems Among Students by using Fuzzy Inference System

Muhammad Faiq Zhafran bin Abdul Gani[1], Muhammad Hani Akmal bin Hefnee[2], Mustafa Nabil bin Abd Malek[3], Siti Nazifah Zainol Abidin[4*], Hazwa Hanim binti Mohamed Hamzah[5]
[1,2,3,4]*Kolej Pengajian Pengkomputeran Informatik & Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, Malaysia*
[5]*Kolej Pengajian Pengkomputeran Informatik & Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, Malaysia*
[*]sitinazifah@uitm.edu.my

Abstract

The prevalence of mental health disorders among university students is a growing global concern, significantly impacting their well-being and academic performance. Many students struggle to recognize their mental health conditions and current assessment methods such as the Depression, Anxiety and Stress Scale-12 (DASS-21) utilized by the counselling unit at UiTM Seremban 3, often lack precision. This study addresses this limitation by aiming to identify the key symptoms affecting student's mental health and developing a more accurate classification model using the Fuzzy Inference System (FIS). The FIS model was constructed through a five-step process which are defining input and output variables, establishing membership functions, formulating a rule-base, conducting fuzzification and performing defuzzification. The results show that the developed FIS model is able to provide more precise value than the DASS-12, with a percentage error of less than 7% for all symptoms analysed. Furthermore, a comparative analysis of students from the three primary faculties at UiTM Seremban 3 reveals that students from Faculty of Sport Science and Recreation (FSR) experience the highest level health issues. This finding suggests that targeted mental health interventions and support programs are especially needed for students in the FSR faculty.

Keywords*:* Mental health problem; Anxiety; Stress; Depression; Fuzzy Inference System

Introduction

Mental health is broadly defined as an individual's overall psychological well-being [1]. It encompasses emotional, cognitive, and social aspects, along with the ability to cope with stress, manage adversity, and function effectively in daily life. Individuals with good mental health exhibit self-confidence, clear thinking, healthy interpersonal relationships, and active contributions to their communities. According to [1], mental health disorders arise from various factors, including genetic predisposition, biochemical imbalances, environmental conditions, and social influences.

Study done by [2] identifies the ten most prevalent mental health disorders, which include anxiety disorders, bipolar affective disorder, depression, dissociative disorders, eating disorders, paranoia, post-traumatic stress disorder (PTSD), psychosis, schizophrenia, and obsessive-compulsive disorder (OCD). However, detecting mental health issues in individuals remains challenging. One common assessment tool is the Depression, Anxiety, and Stress

Scale-21 (DASS-21), which is widely utilized across different countries. The original version of the scale, DASS, comprises 42 items, while its shorter counterpart, DASS-21, consists of 21 items categorized into three interrelated dimensions that measure symptoms of depression, anxiety, and stress [3].

Anxiety and depression are often comorbid, sharing similar symptoms such sleep disturbances, fatigue, muscle tension, and irritability [4]. While anxiety is characterized by excessive and persistent feelings of fear and worry, depression is a mood disorder marked by a sustained period of sadness, loss of interest or pleasure and a significant disruption in daily life [4-5]. Key symptoms of depression, as defined by [5], include persistent low mood, changes in appetite or sleep patterns, feelings of worthlessness, and a lack of energy [5]. These conditions can severely impact a student's academic performance and overall well-being, often requiring treatment through counselling, medication or both [6].

The prevalence of mental health disorders among university students has increased significantly over the past decade. A recent meta-analysis estimated that 33.8% of university students worldwide experience anxiety, while 27.2% exhibit symptoms of depression [7]. These mental health challenges may have long-term consequences, potentially impacting students' professional lives after graduation. Consequently, developing a classification model for assessing student mental health is essential. However, personal factors also play a crucial role in shaping emotions and behaviour [8] and the selection of an appropriate classification model remains challenging due to the wide range of available methods. Several approaches have been explored, including linear regression [9], neuro-fuzzy models [10], convolutional neuro-fuzzy models [11], self-screening and referral systems [12], smartphone sensor-based methods [13] and fuzzy clustering algorithms [14].

Unlike other diagnostic processes, medical diagnosis is inherently complex due to the presence of vagueness, linguistic uncertainty, imprecise measurements, and natural variability [15]. Recent advancements in artificial intelligence (AI) and intelligent computing have facilitated the development of new methodologies that enhance medical decision-making, thereby reducing morbidity, mortality, and economic burdens on healthcare systems [10]. Research conducted by [10] demonstrates that the application of fuzzy systems can improve the assessment of depression severity based on patient case studies. Therefore, this study is motivated to use Fuzzy Inference System to identify symptom of mental health problem.

Related Work

Study done by [10] applied a fuzzy-neural hybrid technique utilizing Mamdani's fuzzy logic controller to assess the severity of depression. Similarly, [9] developed a conceptual framework integrating fuzzy inference systems and linear regression to analyze the impact of various factors on patients' health-related quality of life (HRQL). Collectively, these studies demonstrates the diverse range of fuzzy-based approaches, including clustering algorithms [14] and fuzzy-neural hybrids [16], which have been applied to model complex health-related problems.

Fuzzy Logic Mamdani Method was used by [17] to investigate mental health disorders. The objective of this research was to develop a classification model for identifying characteristics of mental disorders using the Fuzzy Logic Mamdani Method. The proposed system demonstrated high classification accuracy, with an error margin of less than five percent. Additionally, [18] applied the Fuzzy Inference System (FIS) to predict depression risk levels.

Several studies have highlighted that depression is a significant public health concern with a high prevalence among the adult population [10]. Unlike other diagnostic processes, medical diagnosis is inherently complex due to vagueness, linguistic uncertainty, measurement imprecision, and natural variability (Taiwo, 2012). The FIS model has demonstrated an accuracy rate of 94%, with a correlation coefficient of 0.746 between FIS predictions and conventional depression severity scales, indicating its potential to assess depression severity and track its progression. Furthermore, the model effectively predicted 125 cases, exhibiting a strong correlation coefficient ($r = 0.88$, p-value = 0.001) between actual and predicted instances.

Accordingly, studies by [18] focused on forecasting depression risk levels, while [17] employed FIS to classify characteristics of mental disorders. In contrast, the present study explores the classification of students' mental health issues using FIS. The primary distinction between these studies is that while [17] and [18] concentrated solely on mental disorder characteristics, this research integrates their approaches to develop a model capable of determining the severity of symptoms associated with three major mental health conditions: anxiety, stress, and depression.

Methodology

The Figure 1 shows a Fuzzy Inference System (FIS) used in this study. It operates in three main stages in which the first stage is fuzzification that convert a precise numerical input into a fuzzy value representing it as a degree of membership in different categories. Second stage is inferences that uses a knowledge base containing logical rules to process these fuzzy values and make a decision. The last stage is defuzzification in which converts the fuzzy decision back into a single, precise number as the final output of the system.

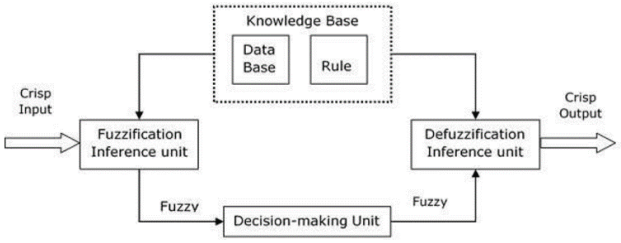


Figure 1. Process in Fuzzy Inference System (FIS)

Data were collected using an online questionnaire, which was distributed to students at UiTM Seremban 3. Although the DASS-21 consists of 21 input variables, this study considered only nine by combining similar questions. The decision to reduce the number of variables was made to limit the number of rules required for the FIS. For each mental health problem, the model was designed with only three input questions. This focused approach resulted in the development of 64 rules for each of the three mental health categories, as the number of rules is calculated by raising the number of linguistic terms to the power of the number of input variables, ensuring a manageable and effective rule base for the FIS. The selected nine variables and their corresponding combined questions from the DASS-21 are detailed in Table 1.

Table 1. Question combination

Mental	Symptoms	Question number in DASS-21
Depression	x_1 :worthless	10, 17, 21
	x_2 :not enthusiatic at all	5, 16
	x_3 :gloomy and sad	3, 13
Anxiety	x_4 :panic and act stupid	9, 15, 20
	x_5 :difficulty breathing	4, 19
	x_6 :mouth felt dry	2, 7
Stress	x_7 :overreact	1, 8, 11, 12
	x_8 :easily offended	6, 14
	x_9 :hard to relax	18

Fuzzification

Table 2 shows fuzzy numbers, scales and membership function for each input variables and output variables using Mamdani method.

Table 2. Fuzzy numbers, linguistic terms and membership functions

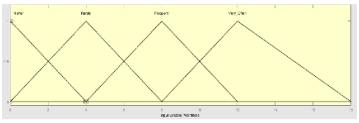
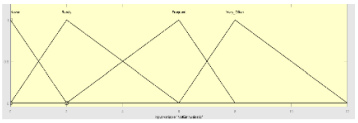
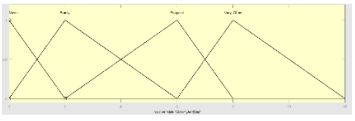
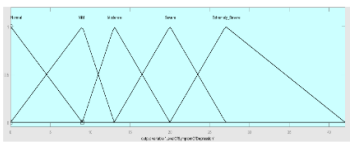
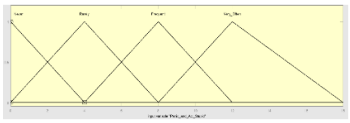
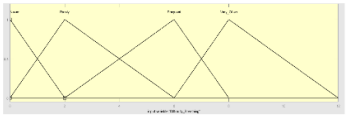
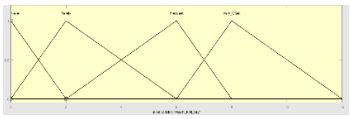
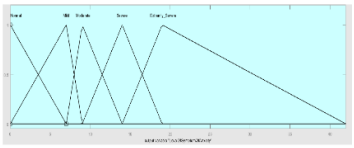
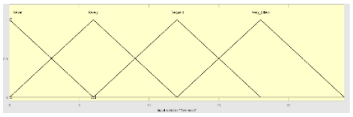
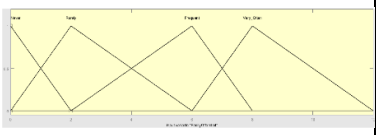
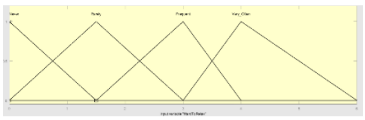
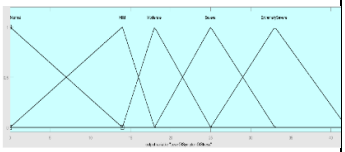
Mental Health Problem	Symptom	Linguistic Terms	Fuzzy numbers	Membership Function
Depression	x_1	Never	0, 0, 4	
		Rarely	0, 4, 8	
		Frequent	4, 8, 12	
		Very often	8, 12, 18	
	x_2	Never	0, 0, 2	
		Rarely	0, 2, 6	
		Frequent	2, 6, 8	
		Very often	6, 8, 12	
	x_3	Never	0, 0, 2	
		Rarely	0, 2, 6	
		Frequent	2, 6, 8	
		Very often	6, 8, 12	
	y_1	Normal	0, 0, 9	
		Mild	0, 9, 13	
		Moderate	9, 13, 20	
		Severe	13, 20, 27	
		Extremely Severe	20, 27, 42	

Table 2 – continues from previous page

Mental Health Problem	Symptom	Linguistic Terms	Fuzzy numbers	Membership Function
Anxiety	x_4	Never	0, 0, 4	
		Rarely	0, 4, 8	
		Frequent	4, 8, 12	
		Very often	8, 12, 18	
	x_5	Never	0, 0, 2	
		Rarely	0, 2, 6	
		Frequent	2, 6, 8	
		Very often	6, 8, 12	
	x_6	Never	0, 0, 2	
		Rarely	0, 2, 6	
		Frequent	2, 6, 8	
		Very often	6, 8, 12	
	y_2	Normal	0, 0, 7	
		Mild	0, 7, 9	
		Moderate	7, 9, 14	
		Severe	9, 14, 19	
		Extremely Severe	14, 19, 42	
Stress	x_7	Never	0, 0, 6	
		Rarely	0, 6, 12	
		Frequent	6, 12, 18	
		Very often	12, 18, 24	
	x_8	Never	0, 0, 2	
		Rarely	0, 2, 6	
		Frequent	2, 6, 8	
		Very often	6, 8, 12	
	x_9	Never	0, 0, 1.5	
		Rarely	0, 1.5, 3	
		Frequent	1.5, 3, 4	
		Very often	3, 4, 6	
	y_3	Normal	0, 0, 14	
		Mild	0, 14, 18	
		Moderate	14, 18, 25	
		Severe	18, 25, 33,	
		Extremely Severe	25, 33, 42	

Evaluate Fuzzy Inferences System

The input and output variables of the membership functions were applied in the FIS for the evaluation process. The output was generated by using IF-THEN rule with condition "IF p is A THEN q is B" where p is the input variable, and q is the output variable and A and B is the linguistic terms. The inference engine uses a system of rules to make decision through the fuzzy "AND" operator and generates a single truth value determines the outcome of the rules. The following is part of the list of rules that was generated.

Rules for Depression

IF x_1 is never AND x_2 is rarely AND x_3 is never, THEN y_1 is normal.
IF x_1 is never AND x_2 is rarely AND x_3 is rarely, THEN y_1 is mild.
IF x_1 is never AND x_2 is very often AND x_3 is rarely, THEN y_1 is moderate.

Rules for Anxiety

IF x_4 is never AND x_5 is never AND x_6 is frequent, THEN y_2 is mild.
IF x_4 is never AND x_5 is frequent AND x_6 is rarely, THEN y_2 is moderate.
IF x_4 is frequent AND x_5 is very often AND x_6 is never, THEN y_2 is severe.

Rules for Stress

IF x_7 is rarely AND x_8 is never AND x_9 is rarely, THEN y_3 is normal.
IF x_7 is frequent AND x_8 is never AND x_9 is never, THEN y_3 is normal.
IF x_7 is very often AND x_8 is very often AND x_9 is frequent, THEN y_3 is extremely severe.

Defuzzification

Defuzzification is the final step in a FIS, converting the system's fuzzy output into crisp values. This study utilizes the Mean of Maximum (MOM) method, a popular approach that calculates the final crisp output by averaging all output values that have the highest membership degree. In this context, the output is represented as Z which is determined by averaging the set of values x_j that share the maximum possibility degree M. The variable k denotes the total number of these output values with the highest degree of possibility [19].

$$Z = \sum_{j=1}^{k} \frac{x_j}{k} x_j \in M \tag{1}$$

Results and Discussion

The result are categorized into three types of mental health problems which are depression, anxiety and stress.

Results for depression

Table 4 shows the output for both DASS-21 and FIS method towards depression which was obtained from the MATLAB software.

Table 3. Result for Depression

Respondents	Faculties	DASS-21		FIS	
		Scoring	Level of Depression	Defuzzification values	Level of Depression
1	KPPIM	24	Severe	20	Severe
2	KPPIM	16	Moderate	14.2	Moderate
3	FSPPP	12	Mild	10.2	Mild
4	KPPIM	16	Moderate	13.8	Mild
5	FSPPP	26	Severe	20	Moderate
6	KPPIM	6	Normal	2.9	Normal
7	FSPPP	26	Severe	29.7	Extremely Severe
8	KPPIM	12	Mild	12.4	Mild
9	FSPPP	12	Mild	10.2	Mild
10	KPPIM	16	Moderate	13.8	Mild
11	FSR	16	Moderate	14	Moderate
12	FSPPP	16	Moderate	13.8	Mild
13	KPPIM	18	Moderate	20	Moderate
14	FSR	28	Extremely Severe	26.6	Extremely Severe
15	KPPIM	24	Severe	21	Severe
16	FSPPP	16	Moderate	13.8	Mild
17	FSR	16	Moderate	13.7	Mild
18	KPPIM	16	Moderate	13.7	Mild
19	FSPPP	24	Severe	26.6	Severe
20	FSR	20	Moderate	20	Moderate
21	KPPIM	12	Mild	9.1	Mild
22	FSPPP	22	Severe	20	Moderate
23	FSR	22	Severe	20	Moderate
24	FSPPP	8	Normal	2.9	Normal

Table 3 – continues from previous page

Respondents	Faculties	DASS-21		FIS	
		Scoring	Level of Depression	Defuzzification values	Level of Depression
25	FSR	28	Extremely Severe	29.7	Extremely Severe
26	FSPPP	26	Severe	26.6	Severe
27	FSR	28	Extremely Severe	26.6	Severe
28	FSR	24	Severe	26.6	Severe
29	FSR	20	Moderate	20	Moderate
30	FSR	12	Mild	12.4	Mild
TOTAL		562	-	524.3	-

Note: Kolej Pen Faculty of Administrative Science and Policy Studies (FSPPP)gajian Pengkomputeran Informatik dan Matematik (KPPIM), Faculty of Sports Science (FSR) and Faculty of Administrative Science and Policy Studies (FSPPP)

Table 3 presents a comparison of the depression classification results from the DASS-21 and the FIS. The results indicate that the FIS model's classification was similar to the DASS-21 for 63% of the cases, demonstrating a high degree of agreement between the two methods. This further supported by the low percentage error of 6.7% for depression, which is well within an acceptable range for a predictive model and shows the FIS model's strong accuracy.

$$\text{Percentage error } = \frac{|524.3 - 562|}{562} \times 100 = 6.7\%$$

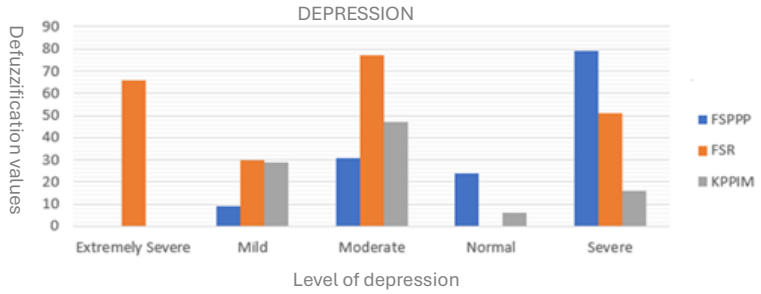


Figure 2. Result for depression between students in UiTM Seremban 3

Based on Figure 2, it shows that Faculty of Sports Science (FSR) has highest level of depression.

Result for Anxiety

Table 4 shows the output for both DASS-21 and FIS method towards anxiety which was obtained from the MATLAB software.

Table 4. Result for Anxiety

Respondents	Faculties	DASS-21		FIS	
		Scoring	Level of Depression	Defuzzification values	Level of Depression
1	KPPIM	18	Severe	14	Moderate
2	KPPIM	14	Moderate	12.6	Moderate
3	FSPPP	20	Extremely Severe	23.5	Extremely Severe
4	KPPIM	18	Severe	23.5	Extremely Severe
5	FSPPP	20	Extremely Severe	25	Extremely Severe
6	KPPIM	14	Moderate	12.6	Moderate
7	FSPPP	12	Moderate	10.2	Moderate
8	KPPIM	12	Moderate	14	Moderate
9	FSPPP	20	Extremely Severe	23.5	Extremely Severe
10	KPPIM	8	Mild	10	Moderate
11	FSR	16	Severe	14	Moderate
12	FSPPP	24	Extremely Severe	23.5	Extremely Severe
13	KPPIM	24	Extremely Severe	26	Extremely Severe
14	FSR	12	Moderate	12.6	Moderate
15	KPPIM	16	Severe	14	Moderate
16	FSPPP	22	Extremely Severe	23.5	Extremely Severe
17	FSR	18	Severe	14	Moderate
18	KPPIM	20	Extremely Severe	23.5	Extremely Severe
19	FSPPP	18	Severe	14	Moderate
20	FSR	26	Extremely Severe	26	Extremely Severe
21	KPPIM	16	Severe	14	Moderate
22	FSPPP	16	Severe	12.6	Moderate
23	FSR	20	Extremely Severe	23.5	Extremely Severe
24	FSPPP	18	Severe	23.5	Extremely Severe

Table 4 – continues from previous page

Respondents	Faculties	DASS-21		FIS	
		Scoring	Level of Depression	Defuzzification values	Level of Depression
25	FSR	24	Extremely Severe	26	Extremely Severe
26	FSPPP	20	Extremely Severe	23.5	Extremely Severe
27	FSR	22	Extremely Severe	25	Extremely Severe
28	FSR	20	Extremely Severe	23.5	Extremely Severe
29	FSR	18	Severe	23.5	Extremely Severe
30	FSR	12	Moderate	12.6	Moderate
TOTAL		538	-	567.7	-

Note: Kolej Pen Faculty of Administrative Science and Policy Studies (FSPPP)gajian Pengkomputeran Informatik dan Matematik (KPPIM), Faculty of Sports Science (FSR) and Faculty of Administrative Science and Policy Studies (FSPPP)

Table 4 shows a 63% similarity rate between the classifications of the FIS and the DASS-21 for anxiety levels. The minimal 5.5% error margin indicates a close alignment between the two methods, confirming the FIS model's strong performance.

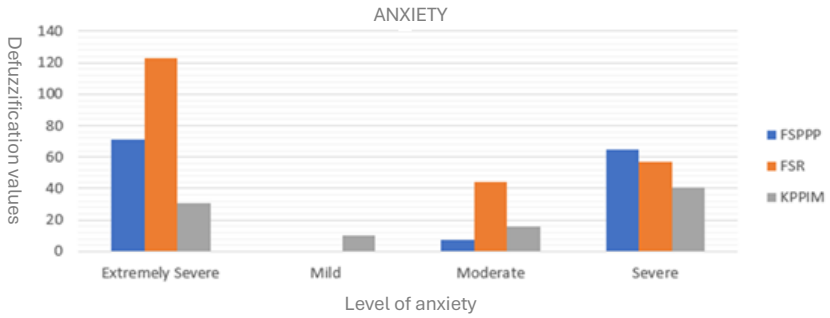


Figure 3. Result for anxiety between students in UiTM Seremban 3

Based on Figure 3, FSR students displayed the highest level of anxiety across the most severe categories, while students from KPPIM generally have the lowest.

Result for Stress

Table 5 shows the output for both DASS-21 and FIS method towards stress which was obtained from the MATLAB software.

Table 5. Result for Stress

Respondents	Faculties	DASS-21		FIS	
		Scoring	Level of Depression	Defuzzification values	Level of Depression
1	KPPIM	10	Normal	5.34	Normal
2	KPPIM	14	Normal	13.1	Normal
3	FSPPP	22	Moderate	25.4	Moderate
4	KPPIM	18	Mild	18.1	Mild
5	FSPPP	12	Normal	13	Normal
6	KPPIM	12	Normal	4.53	Normal
7	FSPPP	24	Moderate	28.6	Severe
8	KPPIM	14	Normal	13.1	Normal
9	FSPPP	24	Moderate	28.6	Severe
10	KPPIM	12	Normal	9.59	Normal
11	FSR	16	Mild	14	Normal
12	FSPPP	14	Normal	14	Normal
13	KPPIM	12	Normal	7.68	Normal
14	FSR	24	Moderate	21	Moderate
15	KPPIM	18	Mild	17.5	Mild
16	FSPPP	14	Normal	8.41	Normal
17	FSR	18	Mild	17.5	Mild
18	KPPIM	14	Normal	14.4	Mild
19	FSPPP	26	Severe	25.4	Moderate
20	FSR	28	Severe	30.8	Severe
21	KPPIM	22	Moderate	18.1	Mild
22	FSPPP	28	Severe	21	Moderate
23	FSR	20	Moderate	18.7	Mild
24	FSPPP	20	Moderate	18.1	Mild
25	FSR	28	Severe	30.8	Severe
26	FSPPP	16	Mild	10.7	Normal
27	FSR	32	Severe	33.4	Severe
28	FSR	22	Moderate	29.1	Severe
29	FSR	26	Severe	28.6	Severe
30	FSR	6	Normal	4.94	Normal
TOTAL		566	-	543.49	-

Note: Kolej Pen Faculty of Administrative Science and Policy Studies (FSPPP)gajian Pengkomputeran Informatik dan Matematik (KPPIM), Faculty of Sports Science (FSR) and Faculty of Administrative Science and Policy Studies (FSPPP)

Table 5 presents the results towards stress, indicating a 63% similarity between the classifications of the FIS and the DASS-21 method. The results of the two methods show only

minor differences, with a low 4% percentage error. This low error rate suggests that FIS model's performance is accurate and acceptable for use.

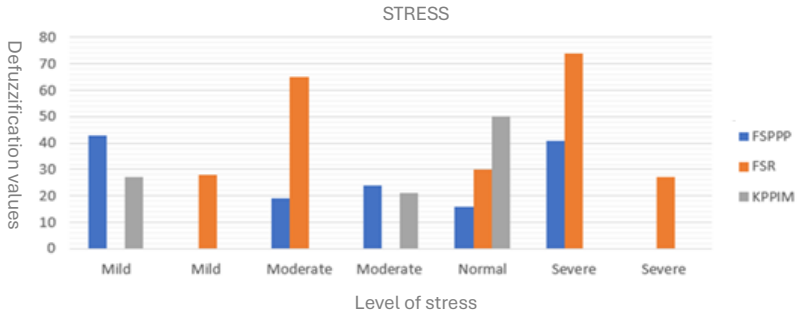


Figure 4. Result for stress between students in UiTM Seremban 3

Based on Figure 4, the results show that students from Kolej Pengajian Pengkomputeran Informatik dan Matematik (KPPIM) have the lowest level of stress compared to students in other faculties. The majority of KPPIM students fall within the normal range on the stress scale, indicating they are less stressed than their peers in FSR and the Faculty of Administrative Science and Policy Studies (FSPPP). The data also indicates that students from FSR experience the highest level of stress.

Conclusion

This study successfully developed a Fuzzy Inference System (FIS) as a classification model for mental health issues, specifically addressing depression, anxiety and stress. The model's development followed a lcear methodology, which included defining input and output variables, establishing membership functions and creating a rule base before undergoing the fuzzification and defuzzification processes.

The results confirm the model's accuracy and reliability, when compared to the DASS-21, the FIS model achieved a high level of similarity, with 63% similarity in classifications for all three mental health categories. furthermore, the model demonstrated a low percentage error of 4% for depression, 5.5% for anxiety and 6.7% for stress indicating its high precision.

A comparative analysis of students from three faculties at UiTM Seremban 3 revealed significant differences. Students from the FSR consistently experienced the highest level of depression, anxiety and stress when compared to their peers in FSPPP and KPPIM.

In conclusion, the FIS model provides accurate tool for assessing student mental health. The findings not only validate its use but also highlight the need for targeted mental health support, particularly for students in the FSR faculty. Further research could explore incorporating all 21 questions from the DASS-21 to enhance the model's complexity and accuracy.

References

[1] Kaligis, F., Ismail, R. I., Wiguna, T., Prasetyo, S., Indriatmi, W., Gunardi, H., Pandia, V., & Magdalena, C. C. (2021). Mental health problems and needs among transitional-age

youth in Indonesia. *International Journal of Environmental Research and Public Health*, 18(8). https://doi.org/10.3390/ijerph18084046

[2] Dryden-Edwards, R., MD. (2023, March 24). Mental Health: Get the Facts on Common Disorders. *MedicineNet.* https://www.medicinenet.com/mental_health/article.htm

[3] Peixoto, E. M., Oliveira, K. D. S., Campos, C. R., Gagnon, J., Zanini, D. S., Nakano, T. D. C., & Bueno, J. M. H. (2021). DASS-21: Assessment of psychological distress through the Bifactor Model and item analysis. *Psico-USF*, 26(4), 757–769. https://doi.org/10.1590/1413- 82712021260413

[4] American Psychological Association (2019, October 28). *What's the difference between stress and* anxiety? Retrieved from https://www.apa.org/topics/stress/anxiety-difference

[5] Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. Emotion, *7*(2), 336–353. https://doi.org/10.1037/1528- 3542.7.2.336

[6] Gale, C., & Oakley-Browne, M. (2000). Anxiety disorder. *BMJ*, *321*(7270), 1204-1207.

[7] Porru, F., Robroek, S. J. W., Bültmann, U., Portoghese, I., Campagna, M., & Burdorf, A. (2021). Mental health among university students: The associations of effort-reward imbalance and overcommitment with psychological distress. *Journal of Affective Disorders, 282*. https://doi.org/10.1016/j.jad.2020.12.183

[8] Paris, M., Hoge, M.A. Burnout in the Mental Health Workforce: A Review. *J Behav Health Serv Res 37*(4), 519–528 (2010). https://doi.org/10.1007/s11414-009-9202-2

[9] Abdullah, L. (2014). Modeling of health-related quality of life using an integrated fuzzy inference system and linear regression. *Procedia Computer Science, 42*, 99–105. https://doi.org/10.1016/j.procs.2014.11.039

[10] Chattopadhyay S, Kaur P, Rabhi F, Acharya R (2012). An automated system to diagnose the severity of adult depression. In *Proceedings of the 2nd IEEE Int. Conf. on Emerging Applications of Information technology*, (pp. 121-124) IEEE. https://doi.org/10.1109/EAIT.2011.17.

[11] Kumar, A., Sangwan, S. R., Arora, A., & Menon, V. G. (2022). Depress-DCNF: A deep convolutional neuro-fuzzy model for detection of depression episodes using IoMT. *Applied Soft Computing*, *122*, 108863. https://doi.org/10.1016/j.asoc.2022.108863

[12] Kim, E. H., Coumar, A., Lober, W. B., & Kim, Y. (2011). Addressing mental health epidemic among university students via web-based, self-screening, and referral system: a preliminary study. *IEEE Transactions on Information Technology in Biomedicine*, *15*(2), 301-307. https://doi.org/10.1109/TITB.2010.2093551

[13] Gunarathne, K. A. A. B., & Jayasena, K. P. N. (2021, December). Monitoring mental health disorder symptoms and behavioral problems of undergraduates in Sri Lanka using smartphone sensors. In *2021 International Conference on Decision Aid Sciences and Application (DASA)* (pp. 544-549). IEEE. https://doi.org/10.1109/DASA53752.2021.9682121

[14] Zhang, Y. (2020, October). Research on college students' psychological crisis early warning method based on fuzzy clustering algorithm. In *2020 International Conference on Computers, Information Processing and Advanced Education (CIPAE)* (pp. 26-30). IEEE. https://doi.org/10.1109/CIPAE50849.2020.00010

[15] Emuoyibofarhe OJ, Taiwo KF (2012). Fuzzy-based system for determining the severity level of knee osteoarthritis, *Int. J. Intell. Sys. Appl.* 9:46-53.

[16] Chattopadhyay, S. (2017). A neuro-fuzzy approach for the diagnosis of depression. *Applied Computing and Informatics, 13*(1), 10–18. https://doi.org/10.1016/j.aci.2014.01.001

[17] Silvana, M., Akbar, R., Derisma, M., Audina and Firdaus (2018). Development of Classification Features of Mental Disorder Characteristics Using The Fuzzy Logic Mamdani Method. In 2018 *International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, Indonesia,* (pp.410-414), IEEE. https://doi.org/10.1109/ICITSI.2018.8696043

[18] Ekong, V. E., Ekong, U. O., Uwadiae, E. E., Abasiubong, F., & Onibere, E. A. (2013). A fuzzy inference system for predicting depression risk levels. *African Journal of Mathematics and Computer Science Research*, *6*(10), 197-204.

[19] Braae M., and Rutherford, D. A., (1978). Fuzzy relations in a control setting. *Kybemetes*, *7*(3), 185-188.

COEFFICIENT BOUND OF A GENERALIZED CLASS OF CLOSE-TO-CONVEX FUNCTIONS USING HERGLOTZ REPRESENTATION THEOREM

Anisa Nur Syahira Lokman[1], Abdullah Yahya[1,*]
[1]*Faculty of Computer and Mathematical Sciences,*
UiTM Cawangan Negeri Sembilan, Kampus Seremban, Malaysia.
[*]abdullahyahya@uitm.edu.my

Abstract

In this article, we introduce a generalized class of close-to-convex functions, $G_K(\alpha,\delta,t)$ *which denoted as* $\mathrm{Re}\left(e^{i\alpha}\frac{f'(z)}{g'(z)}\right)>\delta,(z\in E)$ *where* $|\alpha|\le\pi,\ \cos\alpha>\delta,\ 0\le\delta\le 1,\ -1<t\le 1$ *and* $g'(z)=\frac{1+tz}{1-z}$. *Furthermore, we find the representation theorem and coefficient bound for* $G_K(\alpha,\delta,t)$ *using Herglotz Representation Theorem.*

Keywords*:* Coefficient bound, Close-to-convex, Herglotz representation theorem

Introduction

Complex analysis is the poetry of mathematics which studies a subfield of mathematics that investigates complex numbers, including their functions, derivatives, limits, manipulation, and other mathematical properties. Geometric Function Theory (GFT) is a branch of complex analysis that studies the geometric properties of one-to-one analytic functions. It was started by the German mathematician Ludwig Bieberbach in 1916. According to Duren [1], an analytic function $f(z)$ maps one region to another in the complex plane one-to-one.

Let $\mathbb{C}$ represent complex numbers, and let $f(z)$ be a complex-valued function of the complex variable z. A function f is called univalent in a domain $D\in\mathbb{C}$ if it is one-to-one, meaning that if $z_1,z_2\in D$, $f(z_1)=f(z_2)$, then $z_1=z_2$. (Goodman, [2]). Additionally, a function is called analytic at a point $m\in D$, if it can be differentiated at every point in a neighbourhood around z_0. Let A denoted the class of function of the form

$$f(z)=z+a_2z^n+a_3z^n+\ldots+a_nz^n=z+\sum_{n=2}^{\infty}a_nz^n \tag{1.1}$$

which are analytic in the unit disc, $E=\{z:|z|<1\}$. The function $f(z)$ is also known as normalized univalent function if it satisfies conditions of $f(z)=f'(0)-1=0$ are fixed is denoted by S.

Based on Kaharudin [3], if $f \in S$ is given by (1.1) and $f \in G_K(\alpha,\delta)$, then

$$|a_n| = \frac{2}{n}\left(\frac{1}{2} + A_{\alpha\delta}(n-1)\right), \quad n = 2,3,4...$$

where the functions in this class satisfy the condition

$$\mathrm{Re}\left\{e^{i\alpha}\frac{f'(z)}{g'(z)}\right\} > \delta, \ (z \in E)$$

with $|\alpha| < \pi, \ \cos\alpha > \delta, \quad g'(z) = \frac{1}{1-z}$.

In addition, Yahya, Soh, and Mohamad [4] stated that if $f \in S$ and $f \in G(\alpha,\delta)$ then

$$|a_n| = \frac{2}{n}\left(\frac{1}{2} + \frac{A_{\alpha\delta}(n-1)}{2}\right), \ n = 2,3,4,...$$

where the functions in this class satisfy the condition

$$\mathrm{Re}\left\{e^{i\alpha}\frac{f'(z)}{g'(z)}\right\} > \delta, \ (z \in E)$$

with $|\alpha| < \pi, \ \cos\alpha > \delta, \quad g'(m) = \frac{z}{1-z^2}$.

In the present paper, we focused on the generalized class of close-to-convex functions, $G_K(\alpha,\delta,t)$ which is denoted as S and satisfies the condition

$$\mathrm{Re}\left(e^{i\alpha}\frac{f'(z)}{g'(z)}\right) > \delta, (z \in E) \tag{1.2}$$

where $|\alpha| \le \pi, \ \cos\alpha > \delta, \ 0 \le \delta \le 1, \ -1 < t \le 1$ and $g'(z) = \frac{1+tz}{1-z}$. The main objectives in this study are to find representation theorem and to determine the coefficient bound for the $G_K(\alpha,\delta,t)$ by using Herglotz Representation Theorem.

Preliminaries

To obtain the main result, we applied the Herglotz Representation Theorem to derive the Representation Theorem for $G_K(\alpha,\delta,t)$.

Theorem 1

Let $f \in S$ *and* $f \in G_K(\alpha,\delta,t)$, *then*

$$\frac{e^{i\alpha}\frac{f'(z)(1-z)}{1+tz} - \delta - i\sin\alpha}{\cos\alpha - \delta} = p(z).$$

Proof.

Based on (1.2), let $\frac{f'(z)(1-z)}{1+tz}$ be written as

$$\frac{f'(z)(1-z)}{1+tz} = 1+\sum_{n=1}^{\infty} p_n z^n \,. \tag{2.1}$$

Then, (2.1) can be written as

$$e^{i\alpha}\frac{f'(z)(1-z)}{1+tz} - \delta = 1+\sum_{n=1}^{\infty} p_n z^n .$$

By applying into the relation P, we have

$$e^{i\alpha}\frac{f'(z)(1-z)}{1+tz} - \delta = e^{i\alpha}\left(1+\sum_{n=1}^{\infty} p_n z^n\right) - \delta \,. \tag{2.2}$$

By expanding the equation (2.2)

$$e^{i\alpha}\frac{f'(z)(1-z)}{1+tz} - \delta = e^{i\alpha} + \sum_{n=1}^{\infty} e^{i\alpha} p_n z^n - \delta \,.$$

We can apply the definition of $e^{i\alpha} = \cos\alpha + i\sin\alpha$ into the equation,

$$e^{i\alpha}\frac{f'(z)(1-z)}{1+tz} - \delta = \cos\alpha + i\sin\alpha + \sum_{n=1}^{\infty} e^{i\alpha} p_n z^n - \delta \,.$$

We divide the equation with $\cos\alpha - \delta$ to obtain P ,

$$\frac{e^{i\alpha}\frac{f'(z)(1-z)}{1+tz} - \delta - i\sin\alpha}{\cos\alpha-\delta} = \frac{\cos\alpha - \delta + \sum_{n=1}^{\infty} e^{i\alpha} p_n z^n}{\cos\alpha-\delta} \,.$$

Therefore,

$$\frac{e^{i\alpha}\frac{f'(z)(1-z)}{1+tz} - \delta - i\sin\alpha}{\cos\alpha-\delta} = 1+\frac{\sum_{n=1}^{\infty}\left(e^{i\alpha} p_n\right) z^n}{\cos\alpha-\delta} ,$$

by applying $b_n = \frac{e^{i\alpha} p_n}{\cos\alpha - \delta}$, we have

$$\frac{e^{i\alpha}\frac{f'(z)(1-z)}{1+tz} - \delta - i\sin\alpha}{\cos\alpha-\delta} = 1+\sum_{n=1}^{\infty} b_n z^n \,.$$

We related the function in P with

$$\frac{e^{i\alpha}\frac{f'(z)(1-z)}{1+tz} - \delta - i\sin\alpha}{\cos\alpha-\delta} = p(z). \tag{2.3}$$

So that, $f \in G_K(\alpha,\delta,t)$ if and only if $p(z) \in P$. Based on (2.3), it is noted that $\cos\alpha - \delta$ should always be positive which brings about the condition $\cos\alpha > \delta$ in the definition of the class $G_K(\alpha,\delta,t)$. In addition, by using an approach of Herglotz representation theorem for function in P give a representation function for $G_K(\alpha,\delta,t)$.

Now, we shall prove our main result.

Main Results

In this section, we shall focus on the coefficient bound of $G_K(\alpha,\delta,t)$.

Theorem 2

If $f \in G_K(\alpha,\delta,t)$, *then*

$$|a_n| \le \frac{1}{n}\left(t\left(1-2A_{\alpha\delta}n\right)+1+2A_{\alpha\delta}(n-1)\right), \qquad n=2,3,4,...$$

and equality is attained for each n when f is an extreme point of $G_K(\alpha,\delta,t)$.

Proof.

Given

$$p \in P \Leftrightarrow p(z)=\int_x \frac{1+xz}{1-xz}\,d\mu, \qquad |x|=1,$$

for some probability measure μ on the unit X. Rearranging Theorem 1 to make as the subject,

$$e^{i\alpha}\,\frac{\dfrac{f'(z)}{g'(z)}-\delta-i\sin\alpha}{\cos\alpha-\delta}=p(z).$$

Then,

$$e^{i\alpha}\,\frac{f'(z)}{g'(z)}=p(z)(\cos\alpha-\delta)+\delta+i\sin\alpha,$$

by replacing $\cos\alpha-\delta=A_{\alpha\delta}$, we have

$$e^{i\alpha}f'(z)=g'(z)\left[A_{\alpha\delta}p(z)+\delta+i\sin\alpha\right].$$

Therefore,

$$f'(z)=e^{-i\alpha}\left[g'(z)\left[A_{\alpha\delta}p(z)+\delta+i\sin\alpha\right]\right] \tag{3.1}$$

which implies $A_{\alpha\delta}>0$. From (3.1), we have,

$$f'(z)=e^{-i\alpha}\left[g'(z)\right]\left[(\cos\alpha-\delta)\int_x \frac{1+xz}{1-xz}\,d\mu(x)+\delta+i\sin\alpha\right].$$

Then, follows that

$$f(z)=e^{-i\alpha}\int_0^z\left[g'(\phi)\right]\left[(\cos\alpha-\delta)\int_x\frac{1+x\phi}{1-x\phi}d\mu(x)+\delta+i\sin\alpha\right]d\phi.$$

Thus,

$$f(z)=\int_0^z\left[\int_x\left(\frac{1+xt\phi}{1-x\phi}\right)\left(\frac{1+x\phi\left[e^{-i\alpha}(\cos\alpha-\delta)-e^{-i\alpha}(\delta+i\sin\alpha)\right]}{1-x\phi}\right)d\mu(x)\right]d\phi,$$

and

$$f(z)=\int_0^z\left[\int_x\left(\frac{1+xt\phi}{1-x\phi}\right)\left(\frac{1+x\phi\left[e^{-i2\alpha}-2\delta e^{-i\alpha}\right]}{1-x\phi}\right)d\mu(x)\right]d\phi.$$

Then,

$$f(z)=\int_0^z\left[\int_x\left(\frac{(1+xt\phi)\left[1+x\phi\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)\right]}{(1-x\phi)^2}\right)d\mu(x)\right]d\phi.$$

Now, let $j=e^{-i2\alpha}-2\delta e^{-i\alpha}$,

$$f(z)=\int_0^z\left[\int_x\left(\frac{(1+xt\phi)(1+x\phi j)}{(1-x\phi)^2}\right)d\mu(x)\right]d\phi. \tag{3.2}$$

Based on (3.2), we apply partial fraction to simplify this equation. Therefore

$$f(z)=\int_0^z\left[\int_x\left(jt-\frac{t(2j+1)+j}{1-k}+\frac{t(1+j)+j+1}{(1-k)^2}\right)d\mu(x)\right]d\phi.$$

Therefore, substitute $k=x\phi$ and $j=e^{-i2\alpha}-2\delta e^{-i\alpha}$,

$$\begin{aligned}f(z)=\int_0^z\Bigg[\int_x & t\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)\\ &+\frac{1}{1-x\phi}\Big(-t\left(2\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)+1\right)-\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)\\ &\qquad+\frac{t\left(1+e^{-i2\alpha}-2\delta e^{-i\alpha}\right)+1+e^{-i2\alpha}-2\delta e^{-i\alpha}}{1-x\phi}\Big)d\mu(x)\Bigg]d\phi.\end{aligned}$$

Since,

$$1+e^{-i2\alpha}-2\delta e^{-i2\alpha}=2A_{\alpha\delta}e^{-i\alpha},$$

therefore, we have

$$f(z)=\int_0^z\Bigg[\int_x t\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right) + \frac{1}{1-x\phi}\Big(-2t\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)-t-\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right) + \frac{2tA_{\alpha\delta}e^{-i\alpha}+2A_{\alpha\delta}e^{-i\alpha}}{1-x\phi}\Big)d\mu(x)\Bigg]d\phi \tag{3.3}$$

and

$$f(z)=\int_0^z\Bigg[t\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)\int_x d\mu(x) + \Big(\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)(-2t-1)-t\Big)\int_x\sum_{n=0}^{\infty}(x\phi)^n d\mu(x) + \left(2A_{\alpha\delta}e^{-i\alpha}(t+1)\right)\int_x\sum_{n=0}^{\infty}(n+1)(x\phi)^n d\mu(x)\Bigg]d\phi. \tag{3.4}$$

From (3.4), substitute $n=0$, then

$$\begin{aligned} f(z) &= t\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)+\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)(-2t-1)-t+2A_{\alpha\delta}e^{-i\alpha}(t+1)\\ &= -t\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)-\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)-t+2A_{\alpha\delta}e^{-i\alpha}(t+1)\\ &= 2A_{\alpha\delta}e^{-i\alpha}-\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right) \end{aligned}$$

Substitute $A_{\alpha\delta}=\cos\alpha-\delta$ and $\cos\alpha=\dfrac{e^{i\alpha}+e^{-i\alpha}}{2}$, thus $f(z)=1$.

Then, rewrite equation (3.4) will yield

$$f(z)=\int_0^z\Bigg[1+t\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)\int_x d\mu(x) + \Big(\left(e^{-i2\alpha}-2\delta e^{-i\alpha}\right)(-2t-1)-t\Big)\int_x\sum_{n=1}^{\infty}(x\phi)^n d\mu(x) + \left(2A_{\alpha\delta}e^{-i\alpha}(t+1)\right)\int_x\sum_{n=1}^{\infty}(n+1)(x\phi)^n d\mu(x)\Bigg]d\phi.$$

Integrating with respect to ϕ gives us,

$$f(z) = z\left(1 + t\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)\right) + \left(\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)(-2t-1) - t\right)\int_x \sum_{n=1}^{\infty} (x)^n d\mu(x)\left(\frac{z^{n+1}}{n+1}\right) + \left(2A_{\alpha\delta}e^{-i\alpha}(t+1)\right)\int_x \sum_{n=1}^{\infty}(n+1)(x)^n d\mu(x)\left(\frac{z^{n+1}}{n+1}\right)$$

and

$$f(z) = z\left(1 + t\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)\right) + \left(\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)(-2t-1) - t\right)\int_x \sum_{n=2}^{\infty} (x)^{n-1} d\mu(x)\left(\frac{z^{n}}{n}\right) + \left(2A_{\alpha\delta}e^{-i\alpha}(t+1)\right)\int_x \sum_{n=2}^{\infty}(n)(x)^{n-1} d\mu(x)\left(\frac{z^{n}}{n}\right).$$

Rearranging the equation, we have

$$f(z) = z\left(1 + t\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)\right) + \sum_{n=2}^{\infty}\left[\left(\frac{\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)(-2t-1) - t}{n}\right)\int_x (x)^{n-1} d\mu(x) + \left(\frac{2A_{\alpha\delta}e^{-i\alpha}(t+1)}{n}\right)\int_x (n)(x)^{n-1} d\mu(x)\right] z^n \tag{3.5}$$

and from (1.1), we have that $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$. By comparing (3.5) with (1.1), we have found that

$$a_n = \left[\left(\frac{\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)(-2t-1) - t}{n}\right)\int_x (x)^{n-1} d\mu(x) + \left(\frac{2A_{\alpha\delta}e^{-i\alpha}(t+1)}{n}\right)\int_x (n)(x)^{n-1} d\mu(x)\right] z^n.$$

Rearranging the equation,

$$|a_n| = \frac{\left|-2t\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right) - e^{-i2\alpha} + 2\delta e^{-i\alpha} - t + 2tA_{\alpha\delta}e^{-i\alpha}n + 2A_{\alpha\delta}e^{-i\alpha}n\right|}{n}\int_x \left|(x)^{n-1}\right| d\mu(x).$$

Again since $-e^{-i2\alpha}+2\delta e^{-i\alpha}=1-2A_{\alpha\delta}e^{-i\alpha}$ and $\left|e^{-i\alpha}\right|=1$,

$$\left|a_n\right|=\frac{\left|2t\left(1-2A_{\alpha\delta}e^{-i\alpha}\right)+1-2A_{\alpha\delta}e^{-i\alpha}-t+2tA_{\alpha\delta}e^{-i\alpha}n+2A_{\alpha\delta}e^{-i\alpha}n\right|}{n}\int_x\left|(x)^{n-1}\right|d\mu(x)$$

$$=\frac{\left|t+1+2A_{\alpha\delta}e^{-i\alpha}(n-1)-2tA_{\alpha\delta}e^{-i\alpha}n\right|}{n}\int_x\left|(x)^{n-1}\right|d\mu(x).$$

Thus,

$$\left|a_n\right|=\frac{\left|t\left(1-2A_{\alpha\delta}e^{-i\alpha}n\right)+1+2A_{\alpha\delta}e^{-i\alpha}(n-1)\right|}{n}\int_x\left|(x)^{n-1}\right|d\mu(x)$$

$$\leq\frac{t\left(1-2A_{\alpha\delta}n\right)+1+2A_{\alpha\delta}(n-1)}{n}\int_x\left|(x)^{n-1}\right|d\mu(x)$$

$$=\frac{1}{n}\left(t\left(1-2A_{\alpha\delta}n\right)+1+2A_{\alpha\delta}(n-1)\right)$$

as required. Equality holds when f is an extreme function of $G_K(\alpha,\delta,t)$.

Based on the findings in Theorem 1 and Theorem 2, the representation theorem and coefficient bounds have been established for the class of functions $G_K(\alpha,\delta,t)$.

Conclusion

In conclusion, there are two purposes of this paper, which are to find representation theorem and to determine the coefficient bound for the $G_K(\alpha,\delta,t)$. We believe that we have achieved all the objectives that we have highlighted.

References

[1] P. L. Duren, *Univalent Function*, New York, NY, USA: Springer-Verlag, 1983..

[2] A. W. Goodman, *Univalent Function, Volume 1*. Mariner Publishing Co. Inc. Tampa: University of South Florida, 1983.

[3] N. Kaharudin, "A Close-to-Convex Function", Master Thesis Universiti Teknologi MARA, 2011.

[4] A. Yahya, S. C. Soh, and D. Mohamad, "Coefficient bound of a generalised close-to-convex function". *International Journal of Pure and Applied Mathematics*, *83*(2), 287-293, 2013.

Factors Affecting Rental Prices of Residential Apartments in Kuala Lumpur

Nur Ainaa Azrul[1], Siti Aishah Mohd Shafie[2*], Nornadiah Mohd Razali[3], Az'lina Abdul Hadi[4]
[1,2]*Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan*
[*]ctaishah@uitm.edu.my

Abstract

This study investigates the factors that affect the rental prices of residential apartments in Kuala Lumpur, Malaysia. Thus, as urbanization accelerates and housing demand rises, tenants, landlords, and governments must understand the factors that influence rental costs. Multiple linear regression analysis was used on a dataset of 1335 entries from the Kaggle platform to investigate the relationships between rental prices and six key factors which are the number of rooms, parking spaces, bathrooms, unit size, presence of security, and the location near KTM/LRT. The findings found that all the factors had a significant influence on rental prices of apartments and interestingly, only the number of rooms displayed a negative correlation with rental prices, suggesting niche market preferences. Approximately 33.4% of the total variation in rental prices was explained by the number of rooms, parking spaces, bathrooms, unit size, presence of security, and the location near KTM/LRT. In summary, this research provides useful information for stakeholders to make informed decisions about rental pricing, property investments, and housing policies. In addition, future research should broaden the geographic reach, integrate new factors, and investigate the trends market to better understand apartment rental price dynamics in Kuala Lumpur.

Keywords*:* Affecting, Apartments, Kuala Lumpur, Multiple linear regression, Rental prices

Introduction

The unique qualities of homes determined by age, style, and location are essential since the housing market is not homogeneous [7, 11]. According to [1], rent is considered to be a payment for the use of the building as well as the monthly amount paid to a landlord. Moreover, based on [8], the apartment is a Western housing style with a high population density that was imported and modified to overcome these housing problems such as urban slums and insufficient housing. Apartments are also self-contained housing units that occupy only part of the building and can be rented or purchased. In Malaysia, apartments are provided by both the public and private sectors [8]. In addition, real estate has constantly collected high demand globally, and real estate has grown to be an important sector of the economy with the ability to affect stock market fluctuations and generate unexpected economic events [6]. Apartments constitute an important portion of real estate assets worldwide, because of their amenities, affordability, security, and capacity to house large populations in limited locations, the apartments serve as a viable solution to the housing crisis [18].

Apart from that, due to the inflexible nature of properties and the increase in housing expenses, rental valuation has become increasingly significant, and estimating residential rental values is crucial [24]. However, in highly competitive real estate rental markets, it is necessary to comprehend the variables that could affect rent pricing [21]. On the other hand, residential apartments are solely constructed for domestic usage on "residential" titled land [13]. In East Asia, Malaysia has one of the highest rates of urbanization. As a result, the percentage of people living in cities increased from 26.8% in 1970 to 71% in 2010. It is anticipated that building higher-rise housing will increase the number of units, providing more accommodation and ensuring the continuation of affordable rental schemes [17]. Kuala Lumpur City Hall oversees the management of affordable public housing in Kuala Lumpur, the country's capital. In fact, of the total affordable public housing units available, 37415 units, or 60%, are for rent [10].

According to the News Strait Times, rent increases in the Klang Valley could reach 10 to 20 percent in 2024 as the impact of the COVID-19 movement restrictions fades and demand for real estate increases [25]. Additionally, in Malaysia, housing affordability has been a major problem, particularly for people who live in big cities like Kuala Lumpur and Selangor. Searching for a convenient and affordable home to rent is on everyone's bucket list. However, most people can no longer afford this "affordable housing" due to a sharp rise in house prices [2]. The rental price issue arises when the composition of the rental market consists of landlords who need rent income to pay off their mortgages, and when landlords' rent collections generally show similar upward trends [19]. Some tenants claimed they had to rent such units due to the strategic location close to public transport. Also, the cost to rent these units varies depending on the area and population density in the Klang Valley [15]. Therefore, the objective for this study is to determine the factors that affect rental prices of residential apartments in Kuala Lumpur.

This scope of the study covered several factors, such as the number of rooms, the number of bathrooms, the number of parking space availability, size, security, and proximity to KTM/LRT stations, which may affect the rental prices of residential apartments in Kuala Lumpur. Additionally, the scope of this study will be confined to Kuala Lumpur, the capital city of Malaysia, and specifically targets residential apartments. This focus is due to the high demand of staying in the apartments in the city like Kuala Lumpur. There are also a few limitations to this study. One of the limitations is the findings of this research will be specific to Kuala Lumpur, hence it might not apply to other cities in Malaysia or to different countries with different socio-economic and cultural contexts. Next, the rental market is highly dynamic, with prices fluctuating due to unforeseen factors such as economic crises or natural disasters. The awareness of factors like proximity to public transportation, security, and workplaces can significantly impact a renter's quality of life. Renters can prioritize apartments that offer the best combination of convenience and affordability.

Methodology

Source of Data

The data that has been used for this research was secondary data and the source for the data was from the Kaggle website. However, the data information was originally scraped from the Mudah website which is Malaysia's Largest Marketplace. This study examined 1,335 data points, encompassing variables such as apartment rent, number of rooms and bathrooms, parking availability, unit size, security features, and proximity to KTM/LRT stations. Table 1 below shows the details of all the variables used in this study. All of the variables, apartment rent, rooms, bathroom, parking, and size are numerical variables while, the variables of security and near KTM/LRT are categorical variables.

Table 1: Description of variables

Variable	Description	Level of measurement
Apartment rent	Apartment rent monthly in ringgit Malaysia (RM)	Ratio
Rooms	Number of bedrooms in a unit	Ratio
Bathroom	Number of bathrooms in the unit	Ratio
Parking	Number of parking spaces for the unit	Ratio
Size	Total area of the unit in square feet	Ratio
Security	Presence of security	Nominal (1 = Security, 0 = No security)
Near KTM/LRT	Apartment location near KTM or LRT station	Nominal (1 = Near KTM or LRT, 0 = Not near KTM or LRT)

Theoretical Framework

Figure 1 below shows that the theoretical framework of this study contains one dependent variable, apartment rent, and six independent variables which are rooms, bathrooms, parking, size, security, and near KTM/LRT. It also shows the relationships between the dependent variable and the independent variables.

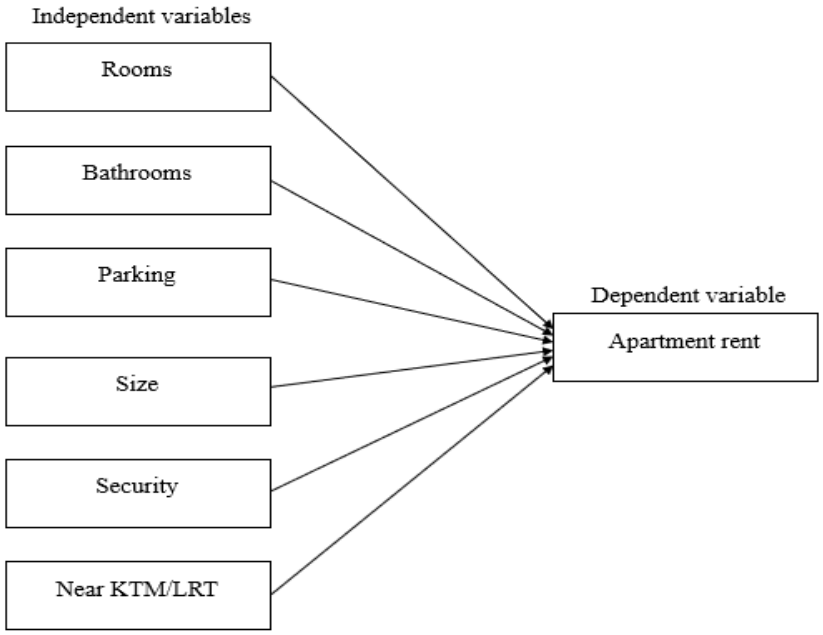


Figure 1: Theoretical Framework

Method of Analysis

Multiple linear regression expands on simple linear regression by adding more than one explanatory variable. The term "linear" is still used in both scenarios because the researchers believe that the response variable is closely related to a linear combination of the explanatory variables [20]. Meanwhile, regression analysis refers to a broad range of statistical methods [5]. Regression is particularly useful for analyzing non-experimental research, and it is frequently more appropriate or simpler to utilize multiple regression to analyze the results of complex quasi-experimental or even experimental research due to the use of dummy variables and contemporary computer software [9]. The general multiple regression model for this study:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \quad (1)$$

where

Y_i is an apartment rent

X_1 is the rooms

X_2 is the parking

X_3 is the bathroom

X_4 is the size

X_5 is the security

X_6 is near KTM/LRT

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ are the parameters of the regression line

ε is the error term

Moreover, a process of transformation towards the variables must be carried out if any assumptions cannot meet the criteria. There are six assumptions which are detecting the linear relationship between each predictor variable and the dependent variables, the error terms are normally distributed, the error variances must be constant (homoscedasticity), the predictors are independent of each other (no multicollinearity), there are no outliers, and the error terms are independent (no autocorrelation) [14].

Results and Discussion

The multiple linear regression analysis was conducted to fulfil the study's objective of identifying the factors influencing residential apartment rental prices in Kuala Lumpur. Despite the removal of all outliers and influential points, the data still failed to meet the necessary assumptions. Consequently, a transformation process was required to ensure compliance with all assumptions.

Transformation of Dependent Variable

In this study, the researcher used the logarithmic transformation of variables. Logarithmically converting variables in a regression model is a frequent method for dealing with non-linear relationships between independent and dependent variables [4]. Furthermore, empirical researchers frequently find the log-transformed approach, in which the dependent variable is natural log-transformed and the unknown coefficients of the explanatory variables in the regression model are estimated using the linear least squares method [3]. After the transformation process, model adequacy checking was conducted again since a new dependent variable was used.

Table 2: Correlations Matrix between Dependent Variables and Independent Variables

		rooms	parking	bathroom	size
ln y	Pearson Correlation	0.073	0.331	0.288	0.455
	Sig.	0.011	0.000	0.000	0.000
	N	1208	1208	1208	1208

Based on table 2, all the p-values are less than the alpha value, 0.05. Hence, the linearity assumption is fulfilled. The p-value for Kolmogorov Smirnov is 0.200 which exceeds the alpha threshold of 0.05, indicating that the error terms are normally distributed. In addition, Park test were conducted to confirm that the homoscedasticity exist.

Table 3: Coefficients of Park test for Independence Variables

Independent variable	Significant Value (p-value)
Rooms	0.915
Parking	0.171
Bathroom	0.211
Size	0.188
Security	0.514
Near KTM/LRT	0.382

As for the statistical method which is the Park test as shown in table 3, all the p-values of the independent variable show that the p-values are greater than the alpha value, 0.05. Therefore, it can be concluded that homoscedasticity exist. The Durbin-Watson value of 1.841 falls within the acceptable range of 1.5 to 2.5, indicating the absence of autocorrelation and confirming the independence of the error terms. In addition, the p-value for the ANOVA, F-test (0.000) is below the alpha threshold of 0.05, indicating the statistical significance of the F-statistic. Consequently, this confirms the overall validity of the model.

Table 4: Coefficient of Independence Variables

Independent variable	Unstandardized coefficients, B	Significant value (p-value)	Tolerance	VIF
Constant	6.417	0.000	-	-
Rooms	-0.0113	0.000	0.763	1.310
Parking	0.080	0.000	0.873	1.145
Bathroom	0.090	0.000	0.712	1.404
Size	0.001	0.000	0.736	1.358
Security	0.138	0.000	0.894	1.119
Near KTM/LRT	0.036	0.009	0.979	1.021

The dataset does not exhibit multicollinearity, as all tolerance (TOL) values exceed 0.2, and all variance inflation factor (VIF) values remain below 10 as shows in table 4. In addition, all the p-values of the independent variable are less than the alpha value, 0.05. This indicates that all independent variables which are rooms, parking, bathroom, size, security, and NearKTM/LRT had a significant relationship with apartment rent. Thus, the results show that all the factors influenced the rental prices of apartments. This means that, the objective had been achieved which to determine the factors affecting rental prices of residential apartments in Kuala Lumpur.

The regression model reveals that all predictor variables, except for the number of rooms, exhibit a positive influence on apartment rent. This indicates that an increase in these predictors

is associated with higher rental prices. Conversely, the negative relationship between the number of rooms and apartment rent suggests that as the number of rooms increases, rental prices tend to decrease. The coefficient of determination, R^2 value of 0.334 implies that 33.4% of the variance in the natural log-transformed apartment rent is attributable to the predictor variables: rooms, parking, bathrooms, size, security, and proximity to KTM/LRT. The remaining 66.6% is influenced by unobserved factors beyond the model's scope. Additionally, the adjusted coefficient of determination, R^2_{adj} value of 0.330, which is marginally lower than the R^2 value, serves as a more robust metric by adjusting for the number of predictors, thereby enhancing the reliability of the model's explanatory power [23].

Conclusion and Recommendations

The analysis confirmed that the selected independent variables, namely the rooms, parking, bathrooms, size, security, and near KTM/LRT, significantly impact rental prices. Interestingly, the number of rooms was negatively correlated with rental prices. This oddity shows that the larger number of rooms in the apartment may serve a niche market or have lower demand than the smaller number of rooms, which is frequently preferred in urban areas. Additionally, apartments with additional security measures had the greatest price premiums, indicating that tenants are willing to pay extra for safety and peace of mind. Meanwhile, the bathrooms and parking moderately positively impacted rental prices. This is consistent with broader market trends that emphasize ease and comfort, as tenants desire additional facilities [16].

The factor of apartment size also demonstrated a positive effect on rental prices. Hence, larger apartments are often regarded to be more desired, yet the incremental price rise for the size of the apartment is much smaller than for other features. Although near KTM/LRT had a positive impact on rental prices, the effect was not too strong. This finding is consistent with global research that shows that the value of location decreases in locations with near KTM/LRT [12]. This research also offers a complete understanding of the factors that influence the rental prices of apartments in Kuala Lumpur. While this emphasizes the significance of the chosen variables, it also demonstrates that a significant portion of the rental price is influenced by other factors such as neighbourhood prestige, building age, or economic conditions [22].

The first recommendation is to analyze temporal trends of apartment rental prices within a year. This is because investigating variations in rental price determinants over time may show changing market preferences, as well as the impact of economic cycles or policy interventions on rental prices of apartments. Other than that, researchers may expand the region and expand the property type of high-rise buildings. Thus, a larger and more diverse region and more property type increases the possibility of detecting significant patterns and insights, improving the overall validity and it would enable policymakers and stakeholders to develop regional-specific strategies. Future research should include more variables in the research such as neighbourhood prestige, building age, proximity to facilities like hospitals, shopping malls, and schools, or other variables that are suitable for this study.

Acknowledgments

We would like to express our gratitude to the Head of College of Computing, Informatics and Mathematics for her advice and encouragement throughout the completion of this research. Besides, we would like to express our gratitude to the Research Ethics Committee, UiTM, for supporting the success of this research. This research was not funded by any grant.

References

[1] Aina, O. C., & Somefun, A. (2007). The effects of facilities provision on rental values of residential properties in ikeja, lagos state. *Journal of Land use and Development Studies*, 3(1), 60–67.

[2] Amit, N., binti Sapiri, H., et al. (2020). Factors affecting housing prices in Malaysia. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(6), 11811–11823.

[3] Anjileli, H., Moftakhari, H., Mazdiyasni, O., Norouzi, H., Ashraf, S., Farahmand, A., Bowler, P., Azarderakhsh, M., Huxman, T. E., & AghaKouchak, A. (2019). Analyzing high-frequency soil respiration using a probabilistic model in a semiarid, mediterranean climate. *Journal of Geophysical Research: Biogeosciences*, 124(3), 509–520. https://doi.org/10.1029/2018JG004640

[4] Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics*, London, Available from: https://kenbenoit.net/assets/courses/me104/logmodels2.pdf

[5] Boldina, I., & Beninger, P. G. (2016). Strengthening statistical usage in marine ecology: Linear regression. *Journal of Experimental Marine Biology and Ecology*, 474, 81– 91. Elsevier BV. https://doi.org/10.1016/j.jembe.2015.09.010.

[6] Cajias, M. (2019). Can a machine understand real estate pricing?–evaluating machine learning approaches with big data (tech. rep.). *European Real Estate Society* (ERES). Available from: https://eres.architexturez.net/doc/oai-eres-id-eres2019-232

[7] Gabe, J., Robinson, S., & Sanderford, A. (2022). Willingness to pay for attributes of location efficiency. *The Journal of Real Estate Finance and Economics*, 65(3), 384–418.

[8] Ju, S. R., Lee, L., & Jeon, S.-Y. (2014). The typologies of block and unit plan in Malaysian apartments focused on middle cost houses supplied by the private sector. *Journal of Asian Architecture and Building Engineering*, 13(2), 397–404. Informa UK Limited. https://doi.org/10.3130/jaabe.13.397

[9] Keith, T. Z. (2019). Multiple regression and beyond: An introduction to multiple regression and structural equation modeling. Routledge. https://doi.org/10.4324/9781315162348

[10] Khair, N., Aini, A. M., & Zyed, Z. A. S. (2022). Public Rental Housing (PRH) operational policies and housing pathways in Kuala Lumpur. *Journal of the Malaysian Institute Planners*, 20(5), 197-208. https://doi.org/10.21837/pm.v20i24.1197

[11] Lecomte, P. (2019). What is smart? A real estate introduction to cities and buildings in the digital era. *Journal of General Management*, 44(3), 128–137. SAGE Publications. https://doi.org/10.1177/0306307018823108

[12] MacCleery, R. (2013). Uli research roundup: The impact of transit on property values. Urban Land Institute. http://uli.org/infrastructure-initiative/uli-research-roundupthe-impactof-transit-on-property-values/

[13] Marzukhi, M. A., Leh, O. L. H., Abdullah, Y. A., Khalid, N. S., Zainol, H., Apet, R. A. A., et al. (2021). The development control process for residential and serviced apartments in shah alam, selangor. *Engineering Journal*, 25(1), 199–209. Faculty of Engineering, Chulalongkorn University. https://doi.org/10.4186/ej.2021.25.1.199

[14] Opreana, A. (2013). A new approach of investment for the future economic policies. *Expert Journal of Economics*, 1(1), 4-12.

[15] Rizan, I. (2023). The more rooms there are, the more profit there is. *Retrieved March 27, 2023*, from https://www.nst.com.my/news/nation/2023/03/893226/more-roomsthere-are-more-profit-there/

[16] Salleh, N. A., Rahman, N. Y. A., Wahab, M. A., & Ahmad, N. (2023). A review on key determinants factors of successful shopping malls. *International Journal of Business and Technology Management*, 5(3), 1–11). https://doi.org/10.55057/ijbtm.2023.5.S3.1

[17] Saari, U. (2022). Kl rental housing - a bad choice for the urban poor? *Retrieved July 21, 2022,* from https://www.malaymail.com/news/what- you- think/2022/07/ 21/kl- rental-housing- a- bad- choice- for- the- urban- poor- urwah- saari/

[18] Samarasinghe, D. A. S. (2020). The housing crisis in australia and new zealand: A comparative analysis through policy lenses. *International Journal of Construction Supply Chain Management*, 10(2), 212–223). https://doi.org/10.14424/ijcscm100220-212-223

[19] Suraya, I. (2022). Renting in malaysia: Can we be 'more-like-germany'. *Khazanah Research Institute*.

[20] Tranmer, M., & Elliot, M. (2020). Multiple linear regression 2nd edition. *The Cathie Marsh Centre for Census and Survey Research (CCSR)*, 5(5), 1–5.

[21] Ullah, F., & Sepasgozar, S. M. (2020). Key factors influencing purchase or rent decisions in smart real estate investments: A system dynamics approach using online forum thread data. *Sustainability*, 12(11), 4382). MDPI AG. https://doi.org/10.3390/su12114382

[22] Uwaezuoke, N. I., Sani, G. S., Igoche, F. O., Akaehomhen, O. N., & Sakariyau, J. K. (2022). Hedonic modelling of residential rental values in ilorin metropolis. *International Journal of Latest Technology in Engineering, Management & Applied Science*, 11(4), 01–09.

[23] Watts, R., Kettner, H., Geerts, D., Gandy, S., Kartner, L., Mertens, L., Timmermann, C., Nour, M. M., Kaelen, M., Nutt, D., et al. (2022). The watts connectedness scale: A new scale for measuring a sense of connectedness to self, others, and world. *Psychopharmacology*, 239(11), 3461–3483). Springer Science and Business Media LLC. https://doi.org/10.1007/s00213-022-06187-5

[24] Wei, C., Fu, M., Wang, L., Yang, H., Tang, F., & Xiong, Y. (2022). The research development of hedonic price model-based real estate appraisal in the era of big data. *Land,* 11(3), 334). MDPI AG. https://doi.org/10.3390/land11030334

[25] Zahiid, S. J. (2024). Real estate agents: Klang valley rental could rise by up to 20pc this year amid recovery from covid-19. *Retrieved February 15, 2024*, from https://www.malaymail.com/news/malaysia/2024/02/15/real-estate-agents-klang-valley-rentalcould-rise-by-up-to-20pc-this-year-amid-recovery-from-covid-19/118111

Forecasting Stock Price in Healthcare Sector by using Geometric Brownian Motion Model

Muhammad Afiq Mukhriz bin Shahruzain[1], Muhammad Nur Hazwan bin Zulfemi[2], Puteri Nurul Atiqah binti Muhamad Nor[3], Siti Nazifah Zainol Abidin[4,*], Aslina binti Omar[5]

[1,2,3,4]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Negeri Sembilan Branch, Seremban Campus, Negeri Sembilan, Malaysia*

[5]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Johor Branch, Pasir Gudang Campus, Johor, Malaysia*

[*]sitinazifah@uitm.edu.my

Abstract

This study addresses the challenge of selecting suitable stocks amid the daily fluctuations and unstable conditions of financial markets. It focuses on the critical need of forecasting tools to predict the future prices that able to assist investor in investment and avoid financial losses. Unlike many studies that focus on the long-term forecasting methods, this study uses Geometric Brownian Motion (GBM) model to forecast short-term prices based investor desire. The model's efficacy is tested using a case study of 10 healthcare stocks, with forecasts generated for two-week and four-week periods. The methodology involves modelling stock price using GBM model, which core component included the rate of return, drift and volatility. The accuracy of the forecasts is validated using the Mean Absolute Percentage Error (MAPE) and percentage increment. The performance of forecasted prices is evaluated against the FBMKLCI market index. Additional financial ratio including Sharpe's and Treynor's indices along with t-test to ascertain risk-adjusted returns and statistical significance. The finding shows that two-weeks forecast period yield the smallest MAPE, indicating that the GBM model provides a highly accurate forecast for short-term investment. A comparison of forecasted ranking with the actual price ranking reveals an 88% correlation, further supporting the model's reliability. Additionally, most of the selected stocks demonstrate a significant correlation with the FBMKLCI. This study contributes to the academic discourse by empirically demonstrating the effectiveness of the GBM model in short-term forecasting in an emerging market, offering valuable insights for investors and financial analysts.

Keywords*:* Geometric Brownian Motion; Forecasting; Healthcare; Stock Prices; Investment

Introduction

Financial market forecasting is a critical needed, particularly in volatile environment of emerging economies like Malaysia [1]. The need for robust forecasted models is paramount, as informed investment decisions are crucial for mitigating risk and capitalizing on market opportunities [1]. In Malaysia, investors can access many of equities traded through Bursa Malaysia, which serves as a central hub for corporate fundraising and public trading [1]. The Financial Times Stock Exchange (FTSE) Bursa Malaysia KLCI also known as FBMKLCI gives investors access to a wide range of complementing indexes and also representing the company's performance While the overall market offers numerous investment avenues, the healthcare sector presents a particulars

compelling area focus. Healthcare sector becomes one of the most popular sectors to invest since Covid-19 and significant growth and continues to attract substantial investor interest [2,3].

Researchers have developed a variety of mathematical models for forecasting stock prices. One of the most famous and commonly used statistical strategies for creating predictions based on past observations is the Autoregressive Integrated Moving Average (ARIMA) [4]. Despite its widespread use in forecasting, this method has significant drawbacks, including seasonality, non-stationarity, and other issues [5]. Artificial Neural Networks (ANN) are one of the most accurate and extensively used for forecasting, pattern recognition, and image processing as a machine learning approach or soft computing methodology [6]. There are other methods also used for forecasting future stock prices [7]. Although various models are widely used, many are better suited for long-term or next-day forecasted and may not adequately capture the short-term forecasted prices that are essential for many investors [8]. The demand for models that can provide short-term forecasts to support rapid decision making is necessary.

This study addresses this gap by proposing and applying the Geometric Brownian Motion (GBM) model to forecast the closing prices of a selected stocks of 10 healthcare stocks listed on Bursa Malaysia. The GBM model is a continuous time stochastic process, well-suited for simulating the random, yet trending, paths of stock prices, making it particularly relevant for short-term analysis [3]. This study aims to provide a forecasting tool for investors seeking to make timely decisions. This study's contribution in application of the GBM model within an emerging market, that offering for short-term investment strategy.

Related Work

Researchers have developed a variety of mathematical models for forecasting share prices. The models include the Markov-Fourier Grey Model (MFGM) by [13], Clustering-Genetic Fuzzy System (CGFS) by [9], High-Order Fuzzy Time-Series Model (HOFTM) by Chen et al., (2008), Moving Average Autoregressive Exogenous (ARX) with combination of Grey System (GS) theory and Rough Set (RS) by [11] and Hidden Markov Model (HMM) by [12].

According to [13] developed a novel approach known as MFGM by combining grey model, Fourier series, and Markov. The MFGM is a strong model that can forecast accurately, but the model is suitable for long term operations. The CGFS model used by [9] is outperforms when compared to HMM, HMM Hybrid, Artificial Neural Network (ANN) and Genetic Algorithms (GA), HMM and Fuzzy Logic Hybrid, ARIMA and ANN. According to [12], HMM model and CGFS model both produce the same outcome, but these models only able to forecast the closing price for the following day. Method like ANN, however, is problematic since it forecasts share prices using fuzzy systems and structures [12]. It also needs some previous information from professionals. Other than that, the method that uses ARX with combination of GS theory and RS [11] is the method that is only suitable for long-term prediction.

Despite what the majority of investors would like, these models are not suitable for short-term investments. According to [1] and [3], Geometric Brownian motion can be used to forecast share prices for one to two weeks of investment in which is more advantageous to investors.

Methodology

Daily data for Malaysia Healthcare sector is used. Ten Healthcare companies which are TOPGLOV, SUPERMX, IHH, HARTA, KOSSAN, AHEALTH, ADVENTA, TMCLIFE, DPHARMA and NOVA are used as case study. The data that has been collected will be used to forecasting future stock prices using the GBM model. At this stage, future stock price forecasting is conducted using the GBM model. In GBM, rate of return is calculated using Eq. 1, where R_i is stock return at time i, S_i is stock price when i, and S_{i-1} is stock price when $i-1$.

$$R_i = \frac{S_i - S_{i-1}}{S_{i-1}}. \tag{1}$$

The next step is to generate the drift and the volatility value [14]. The drift value, μ, also known as the drift coefficient, is defined as the mean of return rate where the price increases as time increases. The drift value is calculated by using Eq. 2, where μ is drift value, R_i is stock return at time i, M is amount of stock return, and dt is time steps.

$$\mu = \frac{1}{Mdt}\sum_{i=1}^{n} R_i \,. \tag{2}$$

Next, the volatility value, also known as fluctuation of the stock price or decreases. In other words, the volatility, σ, is the sample standard deviation of the stock return with the time step. The formula for volatility is as Eq. 3.

$$\sigma = \sqrt{\frac{1}{(n-1)dt}\sum_{i=1}^{n}\left(R_i - \overline{R_i}\right)^2} \,. \tag{3}$$

For the last step, it is to compute the forecasted stock price based on the GBM model [15]. The forecast stock price is predicted using GBM as Eq. 4,

$$S(t) = S(0)e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_i}. \tag{4}$$

where $S(t)$ is future stock value, $S(0)$ is initial stock value, μ is the daily drift using Eq. 6, σ is volatility value using Eq. 1, t is the time step and W_i is the value from probability distribution.

Validate the GBM model

To check the accuracy of the best model of GBM, Mean Absolute Percentage Error (MAPE) are used to validate the forecasting model, percentage increase for increment of stock price and Treynor's index and Sharpe index are used for stock performance.

$$t_v = \frac{r}{\sqrt{\frac{1-r^2}{n-r}}}. \tag{5}$$

where t_v is the t-value, r is the correlation coefficient and n are the number of daily returns. The hypothesis testing will indicate the relationship between the stocks and KLCI.

$$H_0 : \rho = 0 \text{;stock is not related to KLCI}$$
$$H_1 : \rho < 0 \text{;stock is elated to KLCI}$$

The MAPE is a method that is frequently employed to assess forecasting techniques and Eq. 6 shows the MAPE formula.

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|. \tag{6}$$

where A_t is actual stock price at time t, F_t is forecasted stock price at time t, and N is number of stock price data. Table 2 shows the MAPE value accuracy [16].

Table 2. MAPE value accuracy

MAPE	**Accuracy**
< 10%	Highly accurate
10% - 20%	Good forecast
21% - 50%	Reasonable forecast
> 50%	Inaccurate forecast

The forecast accuracy regarding error scale using MAPE has shown that the lower MAPE values is highly accurate in which indicate the forecasting model is accurate. Percentage increase is calculated based on Eq. 7 where I is the percentage increase.

$$I = \frac{A_t - F_t}{A_t} \times 100. \tag{7}$$

The percentage increase used to evaluate which stocks in the Healthcare sector is better in making investment for the future.

Treynor's index measure portfolio performance by dividing portfolio return by the level of portfolio risk [17].

$$T = \frac{R_p - R_F}{\beta}. \tag{8}$$

Treynor's index is calculate based on Eq. 8 where R_p is portfolio return, R_f is risk free rate and β is beta coefficient.

Sharpe index is used to estimate the unit return minus the risk free rate of return relative to the total risk or emphasis the measuring portfolio performance based on non-systematic risk [18].

$$S = \frac{R_p - R_F}{\sigma}. \tag{9}$$

Result and Discussion

The ten stocks is analysed using Eq. 1 until Eq. 4. Table 3 shows the forecast versus actual prices for TOP GLOVE CORPORATION BERHAD by using GBM model.

Table 3. Forecast price and actual price for Top Glove Corporation Berhad

Date	Forecast (RM)	Actual (RM)	Differences
31/05/2022	1.40	1.40	0.00
01/06/2022	1.39	1.37	-0.02
02/06/2022	1.37	1.26	-0.11
03/06/2022	1.25	1.24	-0.01
07/06/2022	1.24	1.25	0.01
08/06/2022	1.28	1.18	-0.10
09/06/2022	1.17	1.22	0.05
10/06/2022	1.05	1.13	0.08
13/06/2022	0.96	1.03	0.07
14/06/2022	1.08	1.03	-0.05
15/06/2022	0.95	1.01	0.06
16/06/2022	1.02	0.98	-0.05
17/06/2022	1.05	1.01	-0.04
20/06/2022	0.96	0.99	0.03
21/06/2022	1.08	1.08	0.00
22/06/2022	1.14	1.01	-0.13
23/06/2022	1.08	1.05	-0.03
24/06/2022	1.14	1.05	-0.09
27/06/2022	1.03	1.04	0.01
28/06/2022	1.02	1.05	0.03
29/06/2022	1.03	1.12	0.09
30/06/2022	1.10	1.04	-0.06

Based on Table 3, it shows comparison between forecasting and actual prices. The differences in forecasting prices and actual prices are quite low. Figure 1 shows a sample of the graphs, illustrates the used the difference between forecast and actual price.

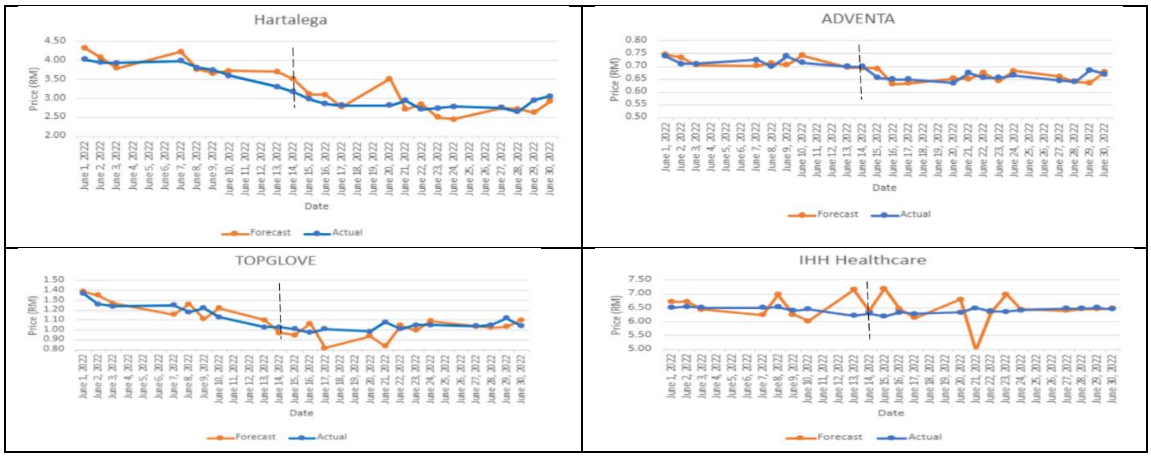


Figure 1. A sample graph of forecasting prices versus actual prices

Based on Figure 1, the graph illustrates that the first two week of investment, the forecast price is quite close to its actual prices as shown in the left part dotted in all the graph. After 2 weeks of investment, the forecast price shows a lot of inconsistency thus the line graph for the forecast will show many fluctuations than the actual one as shown in the right part of the dotted line. This is proven by MAPE value as compared to the result of forecast for 2 weeks and 4 weeks of investment.

Table 4. MAPE value of forecast with actual price for 2 weeks and 4 weeks

Stocks	4 weeks	2 weeks
TOPGLOVE	5.95%	4.40%
IHH	4.14%	3.48%
SUPERMAX	2.82%	2.41%
HARTALEGA	6.13%	5.02%
KOSSAN	7.18%	9.95%
AHEALTH	2.90%	2.32%
ADVENTA	2.22%	2.07%
TMCLIFE	2.49%	1.88%
NOVA	3.86%	1.62%
DPHARMA	3.08%	1.80%

Table 4 shows that the average value of MAPE for two weeks is the smallest.

Table 5. The percentage increment of forecast and actual prices for 10 companies

	Forecast Prices				Actual Prices			
Stocks	31/5/22	14/6/22		Rank	31/5/22	14/6/22		Rank
TOPGLOVE	1.40	1.07	-23%	9	1.40	1.03	-26%	10
IHH	6.70	6.17	-8%	7	6.70	6.30	-6%	6
SUPERMAX	1.05	0.92	-12%	8	1.05	0.91	-13%	8
HARTALEGA	4.20	3.17	-25%	10	4.20	3.17	-25%	9
KOSSAN	1.65	3.21	95%	1	1.65	3.17	92%	1

AHEALTH	2.98	3.32	11%	2	2.98	2.95	-1%	2
ADVENTA	0.75	0.72	-4%	5	0.75	0.70	-7%	7
TMCLIFE	0.53	0.51	-4%	4	0.53	0.51	-4%	4
NOVA	0.75	0.70	-7%	6	0.92	0.89	-3%	3
DPHARMA	1.38	1.39	1%	3	1.38	1.30	-6%	5

Table 5 shows the percentage increment of forecast and actual prices for 10 stocks. The correlation is 88%. Is shown that KOSSAN is the best stock to investment based on the rank. The different between priority stock and percentage increment between actual and forecast prices is not in the large value. However, it is too risky for investor by just referring to percentage increase. Analysis on stock performance using R-square, beta coefficient, Sharpe's index, Treynor's index and correlation analysis is necessary.

Table 6. T-value and correlation for the stocks

Stocks	T-Value	Correlation
TOPGLOVE	1.3595	0.3218
IHH	1.7897	0.4084
SUPERMAX	2.9558	0.5943
HARTALEGA	0.2344	0.0585
KOSSAN	1.0593	0.2560
AHEALTH	0.6496	0.1603
ADVENTA	0.0084	0.0021
TMCLIFE	-1.1100	-0.2674
NOVA	2.3379	0.5047
DPHARMA	-1.4527	-0.3413

Table 6 indicates the t-value and correlation between the company stocks with the KLCI index. Apart from HARTALEGA, ADVENTA, TMCLIFE and DPHARMA whose correlation values are below the significant level of 0.10, it indicates that there is a significant correlation among these stocks and the KLCI index. The highest correlation at 59% is seen between SUPERMAX and KLCI. It suggests that even a slight change in the market index could have an impact on the price of SUPERMAX stock. Additionally, some of the stocks, including TMCLIFE and DPHARMA, have negative values that show a poor relationship with the KLCI index.

Table 7. Stocks performance measured by Treynor's and Sharpe index

Stocks	R-Square	Beta Coefficient	Sharpe Index	Sharpe Rank	Treynor's Index	Treynor Rank
TOPGLOVE	0.1035	-3.8065	-1.9395	5	0.0142	4
IHH	0.1667	-2.8254	-3.2339	9	0.0151	3
SUPERMAX	0.3532	-4.1540	-2.1103	6	0.0114	5
HARTALEGA	0.0034	-2.7428	-0.4386	1	0.0167	2
KOSSAN	0.0655	-1.8380	-0.6657	3	0.0274	1
AHEALTH	0.0257	3.1091	-0.6111	2	-0.0132	7
ADVENTA	0.0000	1.1656	-2.1571	7	-0.0429	8

TMCLIFE	-0.0715	0.4184	-7.2435	10	-0.1077	9
NOVA	0.2547	0.1813	-3.0298	8	-0.2344	10
DPHARMA	-0.1164	4.0311	-1.1839	4	-0.0127	6

Based on Table 7, shows the summary of R-Square, Beta coefficient, Sharpe index and Treynor's index for the 10 stocks. Apart from TMCLIFE and DPHARMA practically all of the stocks have R-Square value above 10. It suggests that the stock index and the KLCI index are correlated. The less sensitive these stocks are, the lower the beta coefficient is. According to the results, the only stocks with beta coefficient values greater than 1 are AHEALTH, ADVENTA and DPHARMA. This indicates that each of these stocks is extremely sensitive to the market. To assess the performance of the stocks, the Sharpe and Treynor's indices are utilized. The performance of each index is contrasted with KLCI index for the time period. The Sharpe index results indicate that, with the exception of the HARTALEGA stock, all of the other stocks underperformed compared to the KLCI index. Meanwhile, the fact that KOSSAN has the highest position on the Treynor's index means that it performs better than other stocks. Based on this finding, it can be seen that KOSSAN outperforms HARTALEGA since it R-Square is larger. This suggests that HARTALEGA's beta is irrelevant for assessing market risk.

Conclusion

This study successfully demonstrates the efficacy of the Geometric Brownian Motion (GBM) model as a valuable tool for short-term financial forecasting in the Bursa Malaysia healthcare sector. The model achieved a high degree of accuracy, with a Mean Absolute Percentage Error (MAPE) of less than 10% for two-week forecast period, confirming its ability to provide reliable forecasted prices. While the four-week forecasts also contributed to the analysis, the two-week period proven to be particularly effective for investors seeking immediate, actionable insights.

The findings provide a clear, data driven framework for making informed investment decisions. A key result is the high degree of reliability demonstrated by the model's output when a comparison of the forecasted price rankings with the actual price rankings revealed an 88% correlation, further supporting the model's utility. Furthermore, most of the selected stocks were found to have a significant correlation with the FBMKLCI, indicating linked to the broader market trend. The analysis of the 10 selected stocks also highlighted a remarkable performance from KOSSAN which showed a 95% increase which means that the price almost doubled up from their actual price at the end of 2 weeks of investment. KOSSAN also show a good result in stock relationship with KLCI index and also stock performance that rank higher than other stocks. This result underscores the potential for significant returns when using GBM for short-term forecasting.

In conclusion, this study validates the GBM model's accuracy for short-term. The findings suggest that GBM can empower investors with a clear advantage, assist in navigate market volatility and identify high-yield opportunities, as demonstrated by the case of KOSSAN. Although stock market investing inherently involves risk, the use of a reliable forecasting model like GBM can significantly increase investor confidence and improve the potential for substantial returns.

References

[1] Abidin, S. N. Z., & Jaffar, M. M. (2014). Forecasting share prices of small size companies in Bursa Malaysia using geometric Brownian motion. Applied Mathematics & Information Sciences, 8(1), 107.

[2] The Kapital (2021, July 30) Sektor Mana Yang Kita Perlu Perhatikan pada tahun 2021?.Retrieved June 10, 2022, from https://thekapital.my/sektor-mana-yang-kita-perlu-perhatikan-pada-tahun-2021

[3] Omar, A., & Jaffar, M. M. (2011, September). Comparative analysis of Geometric Brownian motion model in forecasting FBMHS and FBMKLCI index in Bursa Malaysia. In 2011 IEEE Symposium on Business, Engineering and Industrial Applications (ISBEIA) (pp. 157-161). IEEE.

[4] Meyler, A., Kenny, G., & Quinn, T. (1998). Forecasting Irish inflation using ARIMA models.

[5] Tambi, M. K. (2005). Forecasting exchange rate: A univariate out of sample approach. The IUP Journal of Bank Management, 4(2), 60-74.

[6] Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. Expert Systems with applications, 37(1), 479-489.

[7] Tsai, C.-F., & Wang, S.-P. (2009). Stock price forecasting by Hybrid Machine Learning Techniq? - iaeng.org. Retrieved January 28, 2023, from http://www.iaeng.org/publication/IMECS2009/IMECS2009_pp755-760.pdf

[8] Segal, T. (2022, June 21). What are short-term investments? Investopedia. Retrieved June 27, 2022, from https://www.investopedia.com/terms/s/shorterminvestments.asp

[9] Hadavandi, E., Shanvandi, H., & Ghanbari, A., Integration of Genetic Fuzzy Systems and Artificial Neural Networks for Stock Price Forecasting, Knowledge Based Systems, 23, 2010, pp. 800-808.

[10] Chen, T. L., Cheng, C. H., & Teoh, H. J. (2008). High-order fuzzy time-series based on multi-period adaptation model for forecasting stock markets. Physica A: Statistical Mechanics and its Applications, 387(4), 876-888.

[11] Huang, K. Y., & Jane, C. J. (2009). A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories. Expert systems with applications, 36(3), 5387-5392.

[12] Hassan, M. R., & Nath, B. (2005, September). Stock market forecasting using hidden Markov model: a new approach. In 5th International Conference on Intelligent Systems Design and Applications (ISDA'05) (pp. 192-196). IEEE.

[13] Hsu, Y. T., Liu, M. C., Yeh, J., & Hung, H. F. (2009). Forecasting the turning time of stock market based on Markov–Fourier grey model. Expert Systems with Applications, 36(4), 8597-8603.

[14] Hamzah, S. R., Halul, H., Jeng, A., & Umul Ain'syah Sha'ari. (2021). Forecasting Nestle Stock Price by using Brownian Motion Model during Pandemic Covid-19. Malaysian Journal of Science Health & Technology, 7(2), 58–64. https://doi.org/10.33102/mjosht.v7i2.214

[15] Estember, R. D., John, M., & Maraña, R. (2016). *Forecasting of Stock Prices Using Brownian Motion-Monte Carlo Simulation.*

[16] Ratnayaka, R. K. T., Jianguo, W., & Seneviratna, D. N. (2014, October). Geometric Brownian motion with Ito's lemma approach to evaluate market fluctuations: A case study on Colombo Stock Exchange. In 2014 International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC2014) (pp. 1-6).

[17] Qur'anitasari, Q., Nuzula, N. F., & Darmawan, A. (2021). Critical Analysis of Sharpe,Treynor and Jensen Methods in Analyzing Stock Portfolio Performance LQ-45 Stock Studies. Asia-Pacific Management and Business Application, 8 (2), 89-104.

[18] Pangestuti, I. R. D., Wahyudi, S. & Robiyanto (2017). Performance evaluation of equity mutual funds in Indonesia. Jurnal of Finance and Banking, 21(4): 527-542.

Identifying Factors Shaping Urban Poverty in Malaysia: A Household Perspective

Elis Eliza Rosli[1], Siti Aishah Mohd Shafie[2,*], Az'lina Abdul Hadi[3], Nornadiah Mohd Razali[4]

[1,2,3,4]*Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan*

*ctaishah@uitm.edu.my

Abstract

Urban household poverty refers to the inability of households in urban areas to meet basic needs due to limited income and resources, exacerbated by rapid urbanization and rising living costs. This study investigates the factors contributing to urban household poverty in Malaysia, focusing on demographic, socioeconomic, and regional determinants. Using data from the 2019 Household Expenditure Survey (HES) and logistic regression analysis, the study examines variables such as age, gender, education, occupation, household size, and regional disparities. The findings reveal that lower educational attainment, unstable employment, larger household sizes, and underdeveloped regions are significant contributors to poverty, with female-headed households being particularly vulnerable. These insights provide a foundation for designing targeted, region-specific interventions to reduce poverty, improve socioeconomic resilience, and support sustainable urban development in Malaysia.

Keywords: Determinants, Logistic Regression, Urban household poverty

Introduction

Urban poverty remains one of the most pressing socio-economic issues in Malaysia. Despite the nation's remarkable economic growth and significant strides in reducing poverty, urban households continue to face challenges in meeting basic needs. Urbanization, while a driver of economic progress, has contributed to increasing living costs, income inequality, and disparities in access to essential resources. These issues are compounded by structural and systemic factors such as educational attainment, employment stability, and demographic characteristics, creating a complex web of poverty determinants.

Malaysia's commitment to eradicating poverty is reflected in its alignment with the United Nations Sustainable Development Goals (SDGs), particularly Goal 1, which aims to end poverty in all its forms [2]. However, achieving this goal in urban settings requires a nuanced understanding of the factors contributing to poverty. Urban poverty often manifests differently than rural poverty, encompassing not only income deficits but also challenges such as inadequate housing, limited access to quality education, and unstable employment. This study seeks to fill the knowledge gap by analyzing the demographic, socioeconomic, and regional determinants of urban poverty in Malaysia, providing insights for policymakers and stakeholders to design targeted interventions.

Background of Study

Urban poverty in Malaysia has evolved alongside the country's rapid urbanization. According to the [1], over 75% of the population resides in urban areas, a significant increase from previous decades. This shift has led to improved access to infrastructure and economic opportunities for many, yet it has also highlighted the vulnerabilities of urban households. The urban poor often face high living costs, substandard housing, and limited access to healthcare and education, contributing to a multidimensional experience of poverty. The revised poverty line income (PLI) of RM2,208 per month underscores the financial pressures faced by urban households, with approximately 400,000 households falling below this threshold [1].

Previous studies, such as those by [3] and [4], have identified key factors influencing urban poverty, including household size, education level, gender, and regional disparities. For instance, households with lower educational attainment are more likely to experience poverty due to limited access to stable and well-paying jobs. Similarly, female-headed households often face additional financial burdens due to wage disparities and caregiving responsibilities. Regional disparities further exacerbate poverty, with underdeveloped areas such as Kelantan, Terengganu, Sabah, and Sarawak experiencing higher poverty rates. This study builds on existing literature by using the 2019 Household Expenditure Survey (HES) to examine these factors in detail and provide evidence-based recommendations.

Methodology

To achieve the objectives of this study, a quantitative approach was employed, leveraging data from the 2019 Household Expenditure Survey (HES) provided by the Department of Statistics Malaysia. The dataset comprised information from 12,345 households residing in urban areas in 2019. The analysis focused on identifying significant predictors of urban poverty through logistic regression, a statistical method suitable for modeling binary outcomes. In this context, the dependent variable (Y) represents the poverty status of households, where $Y = 1$ indicates a household below the poverty line and $Y = 0$ otherwise.

The logistic regression model is specified as follows:

$$\text{logit}(P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k \quad (1)$$

where:

- P is the probability of a household being below the poverty line.
- β_0 is the intercept.
- $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients for each of the independent variables.

The independent variables X_k are defined as follows:

- X_1 = Age: Age of the household head.
- X_2 = Gender: Gender of the household head(1=Male, 0 = Female)
- X_3 = Ethnicity: Ethnicity of the household (Malay as benchmark).
  - $X_{3.1}$: 1 if household is Chinese, 0 otherwise;
  - $X_{3.2}$: 1 if household is Indian, 0 otherwise;

  - $X_{3.3}$: 1 if household is others, 0 otherwise.
- X_4 = Marital Status: Marital status of the household head (Married as benchmark).
  - $X_{4.1}$: Never Married, 1 if true, 0 otherwise;
  - $X_{4.2}$: Widow/Widower, 1 if true, 0 otherwise;
  - $X_{4.3}$: Divorced, 1 if true, 0 otherwise;
  - $X_{4.4}$: Separated, 1 if true, 0 otherwise.
- X_5 = Education Level: Highest level of education attained by the household head (Degree/Advance Diploma as benchmark).
  - $X_{5.1}$: 1 if Diploma/Certificate, 0 otherwise;
  - $X_{5.2}$: 1 if STPM, 0 otherwise;
  - $X_{5.3}$: 1 if SPM/SPMV, 0 otherwise;
  - $X_{5.4}$: 1 if PMR/SRP, 0 otherwise;
  - $X_{5.5}$: 1 if No Certificate, 0 otherwise.
- X_6 = Occupation: Occupation type of the household head (Services and Sales Workers as benchmark).
  - $X_{6.1}$: 1 if Manager, 0 otherwise;
  - $X_{6.2}$: 1 if Professional, 0 otherwise;
  - $X_{6.3}$: 1 if Technician and Associate Professionals, 0 otherwise;
  - $X_{6.4}$: 1 if Clerical Support Workers, 0 otherwise;
  - $X_{6.5}$: 1 if Skilled Agricultural, Forestry, and Fishery Workers, 0 otherwise,
  - $X_{6.6}$: 1 if Craft and Related Trades Workers, 0 otherwise;
  - $X_{6.7}$: 1 if Plants and Machine Operators and Assemblers, 0 otherwise;
  - $X_{6.8}$: 1 if Elementary Occupations, 0 otherwise;
  - $X_{6.9}$: 1 if Occupations Not Elsewhere Classified, 0 otherwise.
- X_7 = Region: Region of residence (Different regions as benchmark).
  - $X_{7.1}$: 1 if Western Region (includes Kelantan, Terengganu, Pahang), 0 otherwise;
  - $X_{7.2}$: 1 if Northern Region (includes Kedah, Penang, Perak, Perlis), 0 otherwise;
  - $X_{7.3}$: 1 if Eastern Region (includes Sabah, Sarawak, Labuan), 0 otherwise.
- X_8 = Number of Children: Number of children below 15 years old in the household (Number of child 0 as benchmark).
  - $X_{8.1}$: 1 if there is 1 child, 0 otherwise;
  - $X_{8.2}$: 1 if there are 2 children, 0 otherwise;
  - $X_{8.3}$:1 if there are 3 children, 0 otherwise;
  - …
  - $X_{8.8}$: 1 if there are 8 children, 0 otherwise.
- X_9 = Number of Male Adults: Number of male adults in the household. (Number of male adults 1 as benchmark).

- $X_{9.0}$: 1 if there are no male adults, 0 otherwise;
- $X_{9.2}$: 1 if there are 2 male adults, 0 otherwise;
- $X_{9.3}$: 1 if there are 3 male adults, 0 otherwise;

…

- $X_{9.9}$: 1 if there are 9 male adults, 0 otherwise.

- X_{10} = Number of Female Adults: Number of female adults in the household. (Number of female adults 1 as benchmark).
  - $X_{10.0}$: 1 if there are no female adults, 0 otherwise;
  - $X_{10.2}$: 1 if there are 2 female adults, 0 otherwise;
  - $X_{10.3}$: 1 if there are 3 female adults, 0 otherwise;

  …

  - $X_{10.8}$: 1 if there are 8 female adults, 0 otherwise
- X_{11} = Number of Elderly: Number of elderly persons (aged 55 and above) in the household (Number of elderly person 0 as benchmark).
  - $X_{11.1}$: 1 if there is 1 elderly person, 0 otherwise;
  - $X_{11.2}$: 1 if there are 2 elderly persons, 0 otherwise;
  - $X_{11.3}$: 1 if there are 3 elderly persons, 0 otherwise;
  - $X_{11.4}$: 1 if there are 4 elderly persons, 0 otherwise

There are a number of key assumptions for logistic regression to meet for the result derived to be valid as below:

- Binary Dependent Variable:
  - The dependent variable should be binary, such that containing only two possible values. This is not only a theoretical requirement but also practical. For instance, in a study of poverty, the outcome could be in poverty status, poor or not poor. Satisfaction with this assumption makes sure that the model is applicable to real-world scenarios [14].
- Independence of Observations:
  - The observations must be independent. Hence, the occurrence of an event for one observation should not affect the occurrence of the same event for another observation [15].
- Linearity of Logits;
  - Logistic regression is based on the assumption of a linear relationship between the logit. For example, natural logarithm of the odds-of the dependent and independent variables. This can be checked using the Box-Tid Well test [16].
- No multicollinearity:
  - The independent variables should not be highly inter-correlated. High multicollinearity may distort the estimates and affect the stability of the regression coefficients. Multicollinearity can be detected commonly by using variance inflation factors (VIF) [17]. If the VIF value is greater than 10, it means high multicollinearity.

- Large sample size:
  - Logistic regression requires an adequate sample size. A widely used rule of thumb is that there should be at least 10 samples in the data for every predictor variable in the model. This ensures the stability of the estimate and the possibility to generalize the model [18].
- Absence of Outliers:
  - Outliers can have a huge impact on the estimates in logistic regression. Checking foroutliers and cleaning them helps in ensuring that the model results are not biased [19]. The Mahalanobis distance is the distance of a point from a distribution with variables covariance. It can be used on a dataset to identify and possibly remove outliers.
- Model Fit:
  - In general, the goodness-of-fit of logistic regression can be checked with the Hosmer-Lemeshow test. Its null hypothesis of no difference between observed and model-predicted values, with its *p*-value being good for a fit to a model of value greater than 0.05[14].

The criterion evaluation of logistic regression model, including Omnibus test and Hosmer-Lemeshow test were tested to ensure the validity of the model. The model's goodness-of-fit was evaluated using the Hosmer-Lemeshow test, while the Omnibus Test is used to assess the overall significance of the logistic regression model. The significance of predictors was assessed through Wald statistics and odds ratios while predictive efficiency model measure the overall performance of model. Lastly, both Cox & Snell R-squared and Nagelkerke R-squared offer insights into the proportion of variation explained by the model, allowing researchers to evaluate the predictive power and adequacy of the logistic regression model.

Results and Discussion

Descriptive Analysis

The analysis indicated that 62% of urban households in Malaysia were below the poverty level. In 78% of families, the head of households had only primary or no formal education. Female-headed households made up 32% of those living below the poverty line, highlighting the hardships encountered by single-parent families. Families with more than five members accounted for 45 percent of poor households. The Western Region (Kelantan, Terengganu, Pahang) had the highest proportion of poor households (74%). In comparison, only 13% were in the Northern Region (Kedah, Penang, Perak, Perlis) and 13% in the Eastern Region (Sabah, Sarawak, Labuan). The findings highlight the critical need for targeted initiatives in impoverished regions, as well as educational reforms, to reduce poverty.

Assumptions of Binary Logistic Regression

Firstly, the dependent variable; urban household poverty, is binary in nature, which meets the requirement of logistic regression. Households were classified as either below or above the poverty line, ensuring a clear distinction for analysis.

Secondly, the independence of observations was maintained, as the dataset comprised unique

households without repeated measurements, preventing any potential biases. Additionally, the linearity of the logit was tested using the Box-Tidwell test, which confirmed that the continuous predictor, age of the household head, exhibited a statistically linear relationship with the log odds of poverty after appropriate data transformation and cleaning.

To address potential multicollinearity issues, Variance Inflation Factors (VIFs) were computed for all predictor variables. The results indicated that all VIF values were below the critical threshold of 10, demonstrating that no strong correlation existed among independent variables. Furthermore, the dataset's large sample size of 12,345 observations ensured that the model estimates remained stable and generalizable.

Outliers were identified and removed using Mahalanobis Distance, a statistical measure that detects extreme data points that could influence model accuracy. After eliminating these outliers, the dataset exhibited a more normal distribution, thereby enhancing model reliability. Finally, the model's overall fit was assessed using the Hosmer-Lemeshow test, which resulted in a *p*-value greater than 0.05, confirming that the model adequately fit the data.

Additionally, the Omnibus Test of Model Coefficients validated that at least one predictor significantly contributed to the explanation of urban poverty. The model's predictive efficiency was further reinforced with an overall classification accuracy of 78.65%, which surpassed the acceptable threshold of 60%. This comprehensive validation process ensures that the logistic regression model serves as a robust tool for identifying the determinants of urban poverty in Malaysia.

Binary Logistic Regression Results

The model demonstrated good fit with a Nagelkerke R-squared value of 0.469 as shown in table 1, indicating that approximately 46.9% of the variance in poverty status was explained by the predictors. These results align with prior studies [2, 3], which emphasize the role of education and regional disparities in shaping poverty dynamics.

Table 1: Model Summary

Model	**Cox and Snell R^2**	**Nagelkerke R^2**
Full Model	0.346	0.469

Urban poverty is a multifaceted problem that impairs households' ability to meet necessities. This study examines demographic, socioeconomic, and regional variables to determine the causes of urban household poverty in Malaysia. Significant predictors of poverty were established using the Wald statistic and odds ratios from Table 2. Variables with *p*-values $<$ 0.05 were considered statistically significant, and odds ratios provided information on the degree and direction of their influence.

The age of the household head is a significant factor influencing urban household poverty, with a coefficient of -0.024 and a *p*-value of 0.000. This indicates that as the age of the household head increases, the likelihood of falling into poverty decreases slightly. Age influences poverty because older individuals accumulate more experience and resources, helping to reduce the risk of poverty, while younger individuals often lack assets [3]. Younger

household heads frequently bear the burden of starting a family, managing early career instability, and addressing higher living costs, which makes them more vulnerable to urban poverty [4]. Furthermore, a survey [6] found that older household heads often benefit from secure work, family support systems, or pensions, which reduces their susceptibility to poverty.

Gender is another significant determinant of household poverty levels. Male-headed households are far less likely to fall into poverty, as reflected by a coefficient of -0.554 and a *p*-value of 0.000. This outcome aligns with findings from a study [7], highlighting systemic gender disparities in economic opportunities and wages. Women-headed households in urban centers face a higher likelihood of falling below the poverty line due to income disparities, caregiving responsibilities, and limited access to well-paying occupations. The concept of the "feminization of poverty" underscores the economic instability faced by women, particularly single mothers [8, 9].

Table 2: Parameter Estimates for All Variables

Variables	β	**Wald Statistic**	***p*-value**	**Exp(β)**
X_1	-0.024	64.753	0.000	0.976
X_2	-0.554	27.258	0.000	0.575
$X_{3.1}$	-0.716	137.191	0.000	0.489
$X_{3.2}$	-0.168	2.900	0.089	0.846
$X_{3.3}$	0.725	21.744	0.000	2.065
$X_{4.1}$	0.340	12.102	0.001	1.406
$X_{4.2}$	0.109	0.702	0.402	1.115
$X_{4.3}$	0.198	1.327	0.249	1.219
$X_{4.4}$	-0.181	0.235	0.628	0.835
$X_{5.1}$	0.854	88.706	0.000	2.348
$X_{5.2}$	1.303	67.763	0.000	3.681
$X_{5.3}$	1.738	378.342	0.000	5.685
$X_{5.4}$	2.192	338.459	0.000	8.953
$X_{5.5}$	2.478	396.117	0.000	11.916
$X_{6.1}$	-1.656	233.249	0.000	0.191
$X_{6.2}$	-1.387	186.082	0.000	0.250
$X_{6.3}$	-0.604	49.817	0.000	0.547
$X_{6.4}$	-0.356	9.921	0.002	0.700
$X_{6.5}$	0.745	15.111	0.000	2.107
$X_{6.6}$	0.209	3.842	0.050	1.232
$X_{6.7}$	0.070	0.493	0.483	1.072
$X_{6.8}$	0.642	23.444	0.000	1.900
$X_{6.9}$	0.936	57.011	0.000	2.551

$X_{7.1}$	0.869	104.408	0.000	2.385
$X_{7.2}$	0.810	148.984	0.000	2.247
$X_{7.3}$	0.564	73.933	0.000	1.757
$X_{8.1}$	1.123	220.416	0.000	3.073
$X_{8.2}$	1.623	347.609	0.000	5.069
$X_{8.3}$	2.328	401.708	0.000	10.259
$X_{8.4}$	2.089	157.822	0.000	8.076
$X_{8.5}$	2.260	54.388	0.000	9.587
$X_{8.6}$	2.555	11.263	0.001	12.876
$X_{8.7}$	2.951	6.582	0.010	19.130
$X_{8.8}$	18.186	0.000	1.000	79093289.440
$X_{9.0}$	-0.907	59.532	0.000	0.404
$X_{9.2}$	0.475	47.160	0.000	1.608
$X_{9.3}$	0.403	14.759	0.000	1.496
$X_{9.4}$	0.094	0.256	0.613	1.099
$X_{9.5}$	0.139	0.106	0.744	1.150
$X_{9.6}$	0.644	0.657	0.418	1.904
$X_{9.7}$	-1.792	1.492	0.222	0.167
$X_{9.8}$	-22.687	0.000	0.999	0.000
$X_{9.9}$	-22.789	0.000	0.999	0.000
$X_{10.0}$	-0.994	92.464	0.000	0.370
$X_{10.2}$	0.502	55.040	0.000	1.652
$X_{10.3}$	0.351	10.515	0.001	1.420
$X_{10.4}$	0.487	5.192	0.023	1.627
$X_{10.5}$	-1.031	6.426	0.011	0.357
$X_{10.6}$	1.191	0.690	0.406	3.291
$X_{10.7}$	-21.679	0.000	1.000	0.000
$X_{10.8}$	18.076	0.000	1.000	70862674.876
$X_{11.1}$	0.724	89.993	0.000	2.062
$X_{11.2}$	1.225	88.468	0.000	3.405
$X_{11.3}$	1.459	2.553	0.110	4.302
$X_{11.4}$	21.545	0.000	1.000	2273406269.53

Ethnicity significantly impacts poverty levels. Chinese households are less likely to experience poverty (coefficient of -0.716, *p*-value of 0.000), while households categorized as "Others" are more vulnerable (coefficient of 0.725, *p*-value of 0.000). Certain ethnic groups, such as the Chinese, benefit from better access to education, entrepreneurial networks, and urban economic prospects [10]. Conversely, marginalized groups face systemic disadvantages,

including geographic isolation and limited access to education and healthcare [5].

Education is a critical variable, with lower attainment strongly associated with poverty. For example, heads with no formal education have an odds ratio of 11.916, which means these heads of households are nearly 12 times more likely to experience poverty. Similarly, lower educational levels such as PMR/SRP (coefficient: 2.192, *p*-value: 0.000) and SPM/SPMV (coefficient: 1.738, *p*-value: 0.000) significantly increase the chances of poverty. Education equips individuals with skills and qualifications needed to secure better-paying jobs [11].

Occupation also profoundly influences poverty outcomes. Managerial roles (coefficient: -1.656, *p*-value: 0.000) and professional roles (coefficient: -1.387, *p*-value: 0.000) are associated with a reduced likelihood of poverty, as these jobs provide consistent earnings and additional benefits [3].On the other hand, low-status occupations such as elementary jobs are linked to higher poverty risks.

Regional disparities are highly significant, with households in the Western Region being 2.385 times more likely to fall into poverty compared to others. Similarly, households in the Northern Region (odds ratio: 2.247) and the Eastern Region (odds ratio: 1.757) face significant poverty risks. These disparities stem from underdeveloped infrastructure, limited industrialization, and geographic isolation [5, 10].

Household composition also impacts poverty. Households with one to seven children show significant effects, with odds ratios ranging from 3.073 to 19.130, meaning the likelihood of poverty increases as the number of children rises. Larger families face additional financial burdens, including childcare and education expenses, which exacerbate poverty risks [12]. Similarly, households with one or two elderly members (odds ratios: 2.062 and 3.405) are more vulnerable due to increased healthcare and caregiving costs [13].

Model Efficiency Summary

The logistic regression model's performance in identifying urban poverty outcomes is evaluated using key metrics of sensitivity, specificity, and overall accuracy. Table 3 presents a summary of these efficiency metrics.

The full model demonstrates strong predictive performance, achieving a sensitivity of 86.44%, indicating its effectiveness in correctly identifying many poverty cases. The specificity, at 66.47%, reflects the model's ability to accurately classify non-poverty households, though there is a higher rate of false positives compared to true negatives.

Table 3: Model Efficiency Metrics

Model	Sensitivity (%)	Specificity (%)	Overall Percentage (%)
Full Model	86.44	66.47	78.65

Overall, the model achieves an accuracy of 78.65%, surpassing the acceptable threshold of 60%. This balance between sensitivity and specificity indicates the model's robustness in predicting urban poverty outcomes effectively. The high sensitivity ensures minimal missed poverty cases (false negatives), while the specificity suggests reasonable accuracy in identifying non-poor households. Collectively, these metrics highlight the model's reliability for addressing

urban poverty-related challenges and informing targeted interventions.

Conclusion

This study focuses on the various elements that contribute to urban household poverty in Malaysia, such as demographics, education, occupation, and regional differences. Older household heads are less likely to be poor, whereas female-headed families remain especially vulnerable. Ethnicity has an important effect, with Chinese households less likely to be poor, while excluded populations such as "Others" are more vulnerable.

Lower educational attainment promotes vulnerability; hence education has a big impact on poverty. Similarly, higher-status occupations, such as management and professional positions, have lower poverty risk, whereas low-wage jobs increase it. There are clear regional disparities, with households in less developed regions, such as the Western and Northern regions, suffering greater poverty issues due to inadequate infrastructure and restricted economic opportunity.

Lower education levels make people more vulnerable to poverty. Education plays an important role in reducing poverty. Similarly, having better jobs, like management or professional positions, lowers the risk of poverty. On the other hand, low-paying jobs increase the risk of poverty.

There are also differences between regions. Less developed areas, like the Western and Northern regions, face more poverty because they lack good infrastructure and economic opportunities. Larger families, especially those with many children or elderly members, face more financial struggles. This evidence shows the need for programs to help families.

In short, reducing urban poverty in Malaysia requires specific actions like improving access to education, creating stable jobs, and reducing differences between regions. By focusing on these areas, policymakers can better support families and help reduce poverty in cities.

Acknowledgment

We would like to express our gratitude to the Head of College of Computing, Informatics and Mathematics for her advice and encouragement throughout the completion of this research. Besides, we would like to express our gratitude to the Research Ethics Committee, UiTM, for supporting the success of this research. Our appreciation extends to the Department of Statistics Malaysia (DOSM) for providing access to essential resources and data that were necessary to complete this study. This research was not funded by any grant.

References

[1] Department of Statistics Malaysia. (2020). Malaysia's urbanization rate. *Retrieved from https://www.dosm.gov.my/*

[2] UNESCO. (2021). Education and poverty reduction: The role of education in reducing poverty. *United Nations Educational, Scientific and Cultural Organization.*

[3] Mok, T. Y., Gan, C., & Sanyal, A. (2007). The determinants of urban household poverty in Malaysia. *MODSIM07 Proceedings*, 1124–1130.

[4] Zainalaludin, Z., Saidi, N., Foong, H. F., & Abdullah, S. F. Z. (2021). Relationship between poverty levels and age categories among male and female consumers in Malaysia. *Jurnal*

Pengguna Malaysia, 37, 90–110.

[5] Karimi, M. S., Yusop, Z., & Law, S. (2012). Regional development disparities in Malaysia. *Journal of American Science, 6*.

[6] Mohd, S., Senadjki, A., & Mansor, N. (2016). Trend of poverty among elderly: Evidence from household income surveys. *Journal of Poverty, 22*(2), 89–107.

[7] Zainalaludin, Z., Saidi, N., & Foong, H. F. (2021). Gender and poverty vulnerabilities among urban women in Malaysia. *Jurnal Pengguna Malaysia, 37*, 90–110.

[8] Zainalaludin, Z., & Saidi, N. (2021). Gender and poverty vulnerabilities among urban women in Malaysia. *Jurnal Pengguna Malaysia*. Retrieved from https://jpmjurnal.com/jpm/article/download/74/52.

[9] Karupiah, P., & Gopal, P. S. (2018). Gender socialization and capability deprivation on child urban poverty: Experiences of Malaysian Indian women. *Malaysian Journal of Education and Society*.

[10] Ridzuan, M. I. M., Morshidi, A. B., Zakaria, N. S., Idris, R. Z., & Jambol, D. J. A. (2024). Unraveling the nexus: Exploring Malaysia's research trends in advancing UN SDG 1. *Multidisciplinary Reviews, 8*(2), 2025050.

[11] Mohamed, M. Z., & Xavier, J. A. (2015). Poverty alleviation strategies and new economic model in Malaysia. *International Academic Research Journal of Economics and Finance, 1*(2), 23–32.

[12] Kadir, N. A., Marhaban, S. M., & Rajamanickam, R. (2024). Strengthening child protection in Malaysia: Family-based care and its intersections with the sustainable development goals. *Kurdish Studies, 12*(2), 4158–4176.

[13] Hamid, T. A., & Chai, S. T. (2013). Meeting the needs of older Malaysians: Expansion, diversification and multi-sector collaboration. *Malaysian Journal of Economic Studies, 50*(2), 157–174.

[14] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons.

[15] Menard, S. W. (2010). Logistic regression: From introductory to advanced concepts and applications. Sage.

[16] Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The journal of educational research*, 96(1), 3–14.

[17] Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of interdisciplinary mathematics*, 13(3), 253–26.

[18] Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and cox regression. *American journal of epidemiology*, 165(6), 710–71.

[19] Field, A. (2013). Discovering statistics using ibm spss statistics. Sage.

Chaos in the Waters: Investigating Non-Linear Dynamics Of Rainfall In River Basin

Muhammad Mirza Irfan Bin Mohd Nor[1], Muhamad Luqman Sapini[1,*]
[1]*Centre of Mathematical Sciences, Faculty of Computer and Mathematical Sciences University Teknologi MARA Cawangan Negeri Sembilan Kampus Seremban 70300 Seremban, Negeri Sembilan, Malaysia*
[*]luqman0211@uitm.edu.my

Abstract

Floods are a perennial natural disaster in Malaysia, having significant impacts on urban and rural society. It is essential to recognize the random nature of rainfall to enhance flood prediction and water resource planning. This research explores the evidence of chaos in rainfall patterns in the Ampang (urban) and Hulu Langat (rural) river systems through the 0-1 Test and Lyapunov Exponent (LE) techniques. A ten-year record (2013–2023) from ten chosen rainfall stations was used to identify the evidence of non-linear rainfall time series. The imputation of missing values and reduction in noise through the Single Exponential Smoothing Technique improved the quality of the data, and the chaos test was applied to the improved data. The findings affirm the evidence of chaos in rainfall in urban and rural river catchments, displaying sensitivity to the initial condition and non-linear. Against anticipation, the magnitude of chaos in the urban and rural catchments remained close to each other, indicating the influence of urbanization and land use alteration did not impact the chaos in rainfall. The findings suggest the limitation in conventional hydrological models since they are linear in nature and fail to capture the unpredictability of extreme rainfall events. Through the integration of chaos theory in the study of hydrology, the research offers useful contributions to the improvement of flood forecasting models and disaster readiness. The research underscores the necessity to develop adaptive, non-linear flood forecasting models to accommodate the innate chaos in rainfall patterns. The research findings could support the policymaker and urban planner in the formulation of effective flood protection strategies to build flood resilience to extreme weather events.

Keywords*:* chaos theory, river basin hydrology, 0-1 test, lyapunov exponent, non-linear dynamics

Introduction

Chaos theory is a mathematical theory defining dynamical behaviour as highly sensitive to the condition of dynamic systems. The sensitivity also referred to as the butterfly effect, suggests the outcomes in the system are significantly different depending on the tiny variations in the starting point. Figure 1 illustrates some of the characteristics of the chaos system. Even the random-like nature, the deterministic laws dictate the chaos, and the chaos theory asserts the chaos in the short-term horizon could be predictable, while the unpredictability intensifies in the long term [1]. The chaos theory in meteorology started when Edward Lorenz discovered

variations in the conditions in the air resulting in drastically different patterns in the climate, showcasing the impossibility of meteorology's long-term forecasting [2]. The theory has also found applications in physics, economics, biology, and environmental science, among others, in the use in the research in intricate, structured, and random-like patterns in the systems [3]. The chaos theory in the research in the area of hydrology assists in the research of rainfall patterns, and the patterns are required in the determination of flood risks and water resource planning [4].

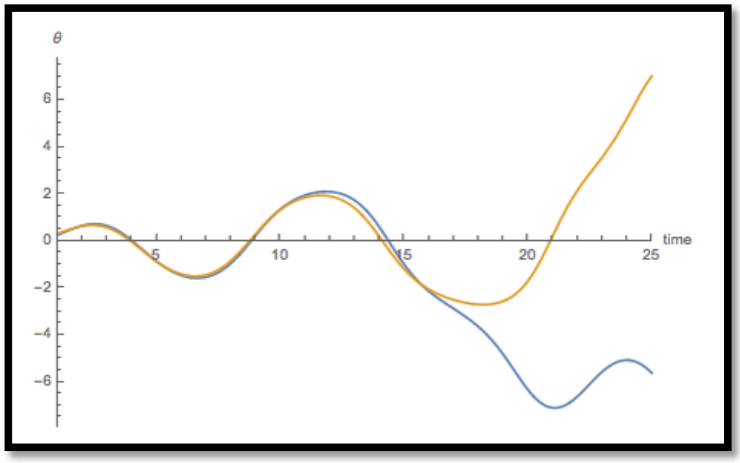


Figure 1: Pendulum that has chaotic system characteristics which are sensitivity to the initial conditions, chaotic motion difficult to forecast and the motion looks random [5].

The use of chaos theory is applied across numerous scientific fields, giving insightful information to systems that have random yet deterministic behavior. In meteorology, chaos theory has helped improve forecasting models by including the inherent unpredictability of atmospheric phenomena (Lu et al., 2010). Within the field of medicine, chaos theory is applied in analyzing physiological signals like heart rate variability and neurological function for the identification of early stages of disorders (Biswas et al., 2018). In finance, it has been used to model stock market volatility and identify patterns within apparently random price changes, assisting investors in making informed decisions and analyzing risk (Peters, 1996). Engineering disciplines use chaos theory to fine-tune control systems to introduce stability in mechanical and electronic processes (Bernardini & Litak, 2016). Environmental applications encompass the study of seismic events, climatic trends, and water cycle processes to more accurately predict natural catastrophes (Domenico et al., 2013). Chaos theory is used in hydrology to study rainfall variability and streamflow dynamics, helping policymakers and urban planners develop efficient flood control strategies (Sapini et al., 2017). Being able to detect chaos in natural phenomena provides a significant edge in predicting and managing environmental extremes.

The atmospheric and hydrological systems possess an intrinsic level of chaos due to their intricate interactions and sensitivity to initial conditions. Meteorologists utilize numerical models in the prediction of meteorological patterns; however, the predictability of the forecasts is inhibited by the non-linearity of atmospheric dynamics (Ahmed, 2023). The butterfly effect, which is a key concept of chaos theory, demonstrates that minute alterations in the weather

patterns can lead to entirely disparate climatic outcomes in the long run (Drew, 2022). For example, a small change in temperature or wind speed will have radical differences in the storm's intensity and path, thus long-term prediction will carry a very high degree of uncertainty (Sapini et al., 2019). River systems and precipitation also exhibit chaotic behavior, in which small differences in atmospheric moisture content will lead to big differences in rainfall and flood occurrence (Sivakumar, 2000). The intrinsic uncertainty involved in flood forecasting models poses enormous challenges since conventional hydrological models are incapable of capturing the non-linear nature of precipitation patterns (Su, 2021). Examining the chaotic dynamics in rainfall patterns can enhance the precision of flood risk analysis, thus allowing concerned authorities to adopt proper disaster management measures (Fauzi et al., 2022). This study investigates the chaotic nature of rainfall patterns to better improve flood prediction models; in Malaysia, incidents of extreme climatic events are quite common.

In Malaysia, flood remains the most devastating climatic disaster that inflicts massive damage to infrastructure, the environment, and human life (Bubeck & Thieken, 2018). The occurrence of flash flood incidents in urban areas such as Ampang, in addition to rural places such as Hulu Langat, highlights the need to expand flood forecasting and management measures (Tella et al., 2023). Flash floods in Kuala Lumpur in January and February 2022 illustrated the unforeseeable nature and magnitude of extreme rainfall events (Tella et al., 2023). Conventional models based on the conventional theory often fail to forecast flood events accurately because rainfall patterns and rivers are inherently non-linear (Sapini et al., 2017). The ineffectiveness of forecasting tools in turn prevents the authorities from undertaking timely disaster responses, and the risks to the people in flood-affected zones increase (Fauzi et al., 2022). This research endeavours to solve this challenge by applying chaos theory to model rainfall dynamics and identifying chaotic patterns in rainfall data from urban and rural river catchments. Through the detection of chaos, this research aims to enhance flood prediction models and support effective water resource management in Malaysia.

This research consists of three objectives, to the effect, detecting the existence of chaos in the river catchments through the 0-1 Test and Lyapunov Exponent, reducing rainfall data noise through the Single Exponential Smoothing Technique, and the comparison between the urban and rural catchment rainfall patterns' behaviours. The novelty in this research lies in the ability to enhance the flood forecasting function, and through the same, inform decision-making by the authorities in flood prevention and control. The realization of the patterns in rainfall chaos helps in the improvement of the models in the hydrological discipline, enhancing the warning mechanisms, and aiding in urban planning to curb flood risks (Sivakumar, 2000). This research also advances the academic discipline by bringing forth evidence in the case of Malaysian rainfall's chaotic nature, supplementing the current research in the area of hydrology. Through the integration between the discipline in chaos theory and the discipline in the study of water in the hydrosphere, the research aims to provide useful contributions to theoretical and applied research in water resource management.

The research methodology in this research comprises the techniques in the acquisition, preprocessing, and chaos detection. Rainfall data are from ten stations, half in Ampang (urban) and half in Hulu Langat (rural) and cover the period 2013-2023. The data are pre-processed, the missing values are imputed by the use of the mean, and the noise in the data is minimized by the use of the Single Exponential Smoothing Technique to increase the precision in detecting chaos. The chaos is detected by the use of the 0-1 Test, to determine if the system follows regular

or chaotic patterns, and the Lyapunov Exponent, to test the sensitivity of the system to the initial condition. The above techniques provide a quantitative measurement of the chaos in the rainfall pattern, and from the measurement, the urban and rural patterns are compared. Through the use of the above techniques, the current research aims to provide a reliable methodology in detecting chaos in rainfall, and in the long run, to increase the precision in flood prediction and water management.

This paper consists of five sections, each having a detailed explanation of the research. The introduction provides an overview of chaos theory, its applications, the problem, the research objectives, and the methodology. The research methodology provides an explanation of the data collection, the preprocessing techniques, and the chaos-detecting techniques used in the research. The result and the discussion provide the findings from the 0-1 Test and Lyapunov Exponent, and the comparison between the urban and rural rivers in the presence of chaos. The conclusion provides the findings, the implications of flood prediction and water management, and the recommendations for future research.

Research Methodology

The research methodology in the current research endeavours to systematically assess the incidence of chaos in rainfall patterns by employing a structured methodology in the acquisition, preprocessing, and processing of the data. This section elaborates the techniques employed to derive rainfall data, preprocess the same to maintain precision, and employ chaos detection algorithms to determine the incidence of the non-linear dynamics in the river systems.

Data Acquisition

This study utilizes rainfall time series data from ten stations selected from the years 2013 to 2023. The study area consists of five stations in the urban area, Ampang, and five stations in the rural area, Hulu Langat. The stations are selected because they are easily flooded and are, consequently, good places to research the random character of rainfall patterns. The ten stations under selection are shown in the list in Table 1.

Table 1: The list and the names of the ten selected stations

Station code	Name of the station
S1	Pusat Penyelidekan at JPS Ampang Selangor
S2	Pemasokan Ampang at Selangor
S3	Bukit Antarabangsa at Selangor
S4	Kg. Berembang at Keramat W. Persekutuan
S5	Gombak Damsite at W, Persekutuan
S6	Stor JPS Hulu Langat at Selangor
S7	Sg. Serai Batu 12 at Hulu Langat
S8	Sg. Raya Bt.9 Hulu Langat at Selangor
S9	Sg. Balak Hulu Langat at Selangor
S10	Balai Polis Batu 14 at Hulu Langat

The data was extracted from the Malaysian Department of Irrigation and Drainage (JPS), which maintains detailed meteorological records across the country. The rainfall measurement each day at each site was recorded and entered into a dataset, forming the basis for further analysis. Since rainfall is highly variable and determined by numerous parameters such as climate variability, land use, and topography, a long-term record was necessary to see patterns and capture chaos (Sapini et al., 2017). The employment of urban and rural sites allows comparative investigation into how the activity of people and the natural world dictates rainfall patterns and flood risks.

Data Preprocessing

Before chaos detection, the data received from the sampling sites underwent preprocessing to enhance accuracy and dependability. Since raw hydrological datasets are prone to having missing values and noise from instrument errors, imputation and reduction techniques in the data were carried out (Bhandari, 2022).

Missing data, also known as missing values, occur in a dataset where no available data is stored for specific variables or participants. A variety of reasons, including the deletion of files, dysfunctional equipment, and incorrect entry, could result in the disappearance of data (Bhandari, 2022). The missing values were addressed through the use of mean imputation, a commonly applied statistical technique that substitutes missing entries with the average of available data points. This method ensures consistency in time series analysis while maintaining data integrity. Eq. 1 is the formula for mean amputation, $\bar{X}$:

$$\bar{X} = \frac{\sum X_i}{n} \tag{1}$$

where $\sum X_i$ is the sum of available data value and n is the number of summed data.

Additionally, noise reduction was also accomplished by applying the Single Exponential Smoothing Technique (Eq. 2), smoothing the variability in the time series by assigning higher weights to the most up-to-date observations and reducing the impact of short-term fluctuations. The smoothing parameter (α) was determined at 0.3, based on the outcomes in other research, in balancing responsiveness and stability in the employment of hydrological data.

$$f_{t+m} = \alpha y_t + (1 - \alpha) f_t \tag{2}$$

where f_{t+m} is the Single Exponential Smoothed value in period $t + m$, for $m = 1, 2, 3, ...,$.

α is the unknown smoothing parameter to be set from 0 to 1. y_t is the actual value in time, t. f_t is the forecast or smoothed value for time, t. The preprocessed data was then divided into original and smoothed datasets, enabling comparisons between raw and refined data when detecting chaotic patterns.

Chaos Detection Techniques

To analyze the presence of chaos in rainfall patterns, two widely recognized methods were employed: the 0-1 test and the Lyapunov Exponent (LE).

The 0-1 Test provides a good and simple tool to distinguish between regular and chaotic dynamics in deterministic systems. It calculates a statistical value, denoted by K, whose values approaching 1 describe the case of chaos, and whose values approaching 0 describe the case of regularity (Gottwald & Melbourne, 2016). The equation of the 0-1 test is shown in Eq. 3 to Eq. 6.

1) For $c \in (0, \pi)$, the following variable are calculated:

$$p_C(n) = \sum_{j=1}^{n} x(j)cos\ (jc) \tag{3}$$

$p_C(n)$ is test trajectory, x_j is the data point, $\cos(jc)$ a periodic weighting function applied to each data point and c is a constant from 0 to 2π to get the median of K.

2) The mean square displacement M_c will be calculated:

$$M_c = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} [p_c(j+n) - p_c(j)]^2 \tag{4}$$

where $[p_c(j+n) - p_c(j)]$ is the difference in trajectory position at times $j+n$ and j.
$[p_c(j+n) - p_c(j)]^2$ is the squared displacement, the spread of the trajectory.
$\lim_{N\to\infty}$ to make sure that the mean square displacement is evaluated for the entire data set.

3) The next step is to estimate the asymptotic growth rate, K_c:

$$K_c = \lim_{n\to\infty} \frac{\log M_c(n)}{\log(n)} \tag{5}$$

$\log M_c(n)$ is the logarithm of MSD, $\log(n)$ is the logarithm of time, $n\ and\ \frac{\log M_c(n)}{\log(n)}$ is the growth rate of the MSD over time

4) Repeat step 1, 2, and 3 for numerous values of c, different randomly selected where $c \in (0, \pi)$.

$$K = median\ (K_c) \tag{6}$$

If the value of $K \approx 0$, it indicates normal dynamics, whereas if the value of $K \approx 1$, it signifies chaotic dynamics.

The second technique, Lyapunov Exponent (LE), computes the rate at which small disturbances in the system grow over time. A positive LE value confirms the presence of chaos, while a negative or zero value indicates a stable or periodic system (Fauzi et al., 2022). This method is particularly useful in hydrological studies as it quantifies sensitivity to initial conditions, a key characteristic of chaotic systems (Sapini et al., 2017). The equation for the Lyapunov Exponent (LE) is presented in Eq. 7 to Eq.8 (Fauzi et al., 2022). The value of the

Largest Lyapunov Exponent (LLE) can be determined by computing the mean logarithmic rate of separation,

$$L_k = \frac{1}{2(N-k-m+1)} \sum_{n=m}^{N-k} log \sum_{j=0}^{m-1} (x_{l-j+k} - x_{n-j+k})^2 \quad (7)$$

where m is a value that represent the value of the embedding dimension, N is the total number of data available, k is the number of intervals, x_l is the value of the signal at time, $l$and L_k is the local divergence at time step, k.

To prevent inaccuracies in the results, the extremely small and large values of k should be excluded. This is because small values of k may contain noise and round-off errors in the separation calculations, as they are not aligned along the direction of greatest expansion. In the meantime, the high values of k, generate a separation that eventually gets close to the attractor's size. The Lyapunov Exponent is defined as Eq.8:

$$\lambda_1 = \frac{dL_k}{dk} \quad (8)$$

If the value λ_1 is greater than zero, the system exhibits chaos and instability. The test was conducted on both the original and smoothed rainfall datasets to assess the impact of noise reduction on chaos detection. By utilizing these two techniques, the study aims to provide a comprehensive evaluation of chaotic rainfall patterns and their implications for flood forecasting.

Results and Discussions

This chapter presents the results and discussions derived from the study on the analysis of rainfall data through the application of chaos detection algorithms. The results are organized into three broad categories: the performance of data preprocessing, chaos detection using the 0-1 Test and Lyapunov Exponent, and comparative analysis of chaotic dynamics in urban and rural river systems.

Effectiveness of Data Preprocessing

Data preprocessing is essential to ensure the quality and reliability of results from chaos detection methods. Mean imputation was successful in addressing the missing value problems within the dataset, ensuring data integrity and facilitating precise time series analysis. Table 2 shows a sample of rainfall data with the application of mean imputation. The missing value denoted by NaN was imputed with the new mean imputation value of 5.3733.

Table 2: The sample of raw rainfall data for the ten selected stations with calculated mean imputation

Date	Raw data	Imputation data
01/07/2018	6.0	0.4
.	.	.
10/07/2018	**NaN**	5.3733
11/07/2018	**NaN**	5.3733
12/07/2018	**NaN**	5.3733

Also, the application of the Single Exponential Smoothing Technique was successful in removing noise from rainfall data, thus making trends more apparent without altering the integrity of the original dataset. As observed in Table 3, the smoothed dataset was less fluctuating compared to the raw data, yielding a cleaner dataset for chaos detection. Nevertheless, although noise reduction enhances data quality, it also carries with it a risk of diminishing some chaotic nature. The analysis of raw and smoothed outcomes indicates that, while noise reduction results in record stability for rainfall, it does not remove the intrinsic chaotic dynamics of river systems (Sapini et al., 2017). The role of preprocessing in hydrologic research is therefore highlighted, specifically the proper detection of chaos in time series data sets.

Table 3: The sample of original and smoothed data for S1

Date	Original Rainfall	Smoothed Rainfall
01/01/2013	0.0	0.0
02/01/2013	0.0	0.0
03/01/2013	0.0	0.0
04/01/2013	0.5	0.15
05/01/2013	15.5	4.755
06/01/2013	0.0	3.3285
07/01/2013	0.0	2.3299
08/01/2013	0.0	1.6309
09/01/2013	0.0	1.1416
. . .	. . .	. . .
30/01/2013	1.0	2.5417

Detection of Chaos in Rainfall Data

The existence of chaos in the rainfall time series was verified by the use of the 0-1 Test coupled with the Lyapunov Exponent (LE) test. As evident in Figure 2, the 0-1 Test outcomes revealed that K values for all ten stations were near 1, hence verifying the chaotic nature of both urban and rural river systems. This validates the assertion that rainfall dynamics are non-linear,

rendering extreme events such as flash floods difficult to predict (Gottwald & Melbourne, 2016).

Likewise, the Lyapunov Exponent (LE) test also provided additional evidence for the existence of chaotic dynamics because all observation stations gave positive LE values that provide evidence of sensitivity to initial conditions, as illustrated in Figure 3. Such results corroborate the hypothesis that precipitation regimes of river catchments are not in a purely deterministic or periodic form, but rather in the features of chaotic systems. The existence of chaos implies that traditional flood forecasting models, which tend to rely on linear relationships, are inadequate to describe the complexity of rainfall dynamics, and more sophisticated, non-linear hydrological models are required (Sapini et al., 2019).

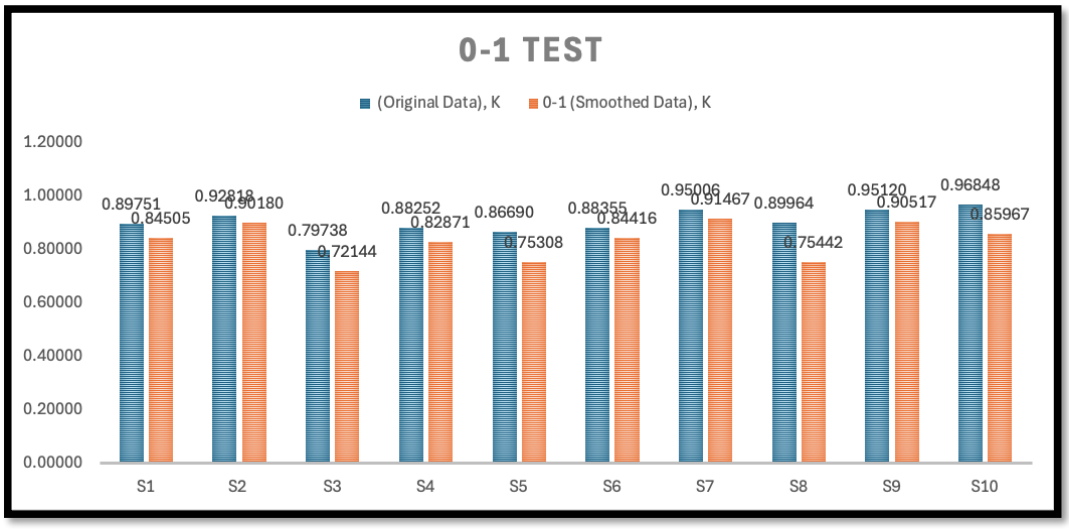


Figure 2. *K* values from 0-1 test for all ten stations

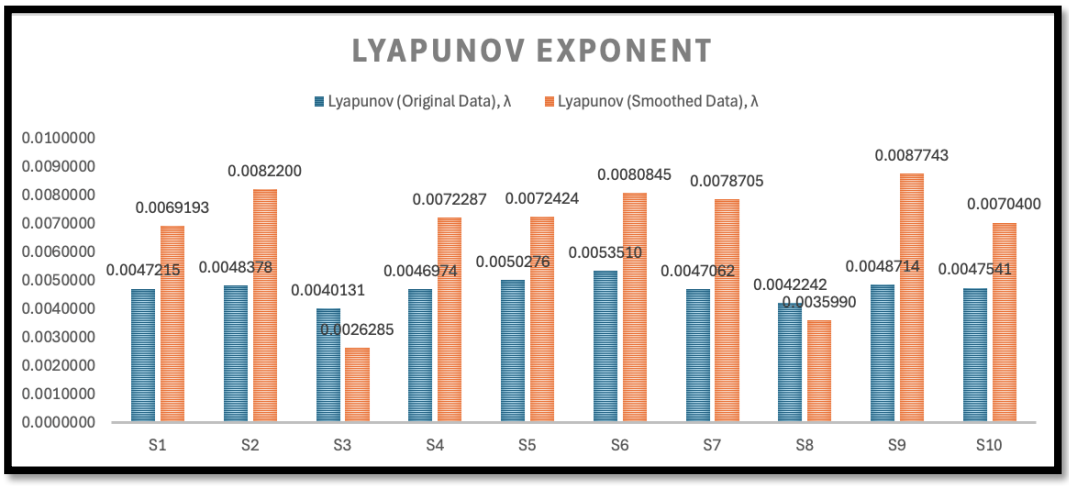


Figure 3. Lyapunov Exponents values for all ten stations

Comparison of Chaos Between Urban and Rural River Systems

This research essentially attempted to test how much chaos was present in rural and urban river basins. Results from 0-1 Test and Lyapunov Exponent (LE) analyses suggest that both urban (Ampang) and rural (Hulu Langat) river systems exhibit chaotic rainfall of comparable magnitude. The K values of the 0-1 Test (Figure 2) and LE values (Figure 3) of all the stations are highly correlated, indicating that the chaotic behavior of precipitation is not affected by the urban or rural character of the environment. The results suggest that the rainfall process is inherently chaotic, independent of infrastructure, anthropogenic modification, or land use.

Earlier conjectures implied that rural locations might exhibit stronger chaotic behavior due to the lack of human impact and inherent hydrological variability. The findings, however, indicate that urban locations also preserve high chaos levels. This is due to the intricate interaction between natural precipitation processes and engineered drainage infrastructure, which can contribute to heightened uncertainty instead of forming stable hydrological regimes (Sapini et al., 2017). Urban infrastructure, including storm drainage systems and flood controls, modifies the hydrodynamics of flow without removing the intrinsic chaotic nature inherent in precipitation. Rural river systems, despite being relatively subject to fewer anthropogenic modifications, still retain the non-linear dynamics owing to the intrinsic complexities in watershed hydrology and effects of climatic factors (Sivakumar, 2000).

The same level of chaos engaged in both settings underscores the intricacy in forecasting floods and managing water resources in varying landscapes. The research refers to the use of chaos theory-based flood forecast models in all settings, as opposed to presuming that urban and rural settings require fundamentally different prediction mechanisms. Given the fact that chaos prevails in both cases, the application of real-time monitoring systems coupled with non-linear prediction methods may enhance the quality of flood risk analysis (Fauzi et al., 2022). All these findings point to the necessity of adaptive responses to flooding according to the intrinsic uncertainty of rainfall features, so that both urban and rural settlements are adequately prepared for such extreme events.

Conclusion and Recommendations

Using the 0-1 Test and Lyapunov Exponent (LE) methods, this study investigates chaos of rainfall patterns in the Ampang and Hulu Langat river basins in urban and rural settings, respectively. Results confirm that rain dynamics in both contexts are chaotic with respect to initial conditions and non-linear. The study revealed that the level of chaos in river systems remains nearly the same in both urban and rural areas. This suggests that while urbanization alters water flow and flooding patterns, it does not substantially change the chaotic nature of rainfall itself. While we were able to get a more accurate data quality based on the application of the Single Exponential Smoothing Technique, the stochastic nature of rainfall continued to be a challenging factor. These findings underscore the limitations of linear assumption hydrological models, adding force to demands for non-linear flood prediction approaches considering chaotic rainfall variability. By incorporating chaos theory into hydrological studies, this article comes with opportune insights for improving flood prediction models and disaster preparedness interventions, particularly in Malaysia's flood-prone zones.

Since rainfall has an inherent uncertainty, flood forecasting models must incorporate non-linear approaches to enhance predictability. That both urban and rural areas are characterized by the same chaotic behavior indicates that land use change and development do not eliminate the intrinsic unpredictability of rain. This justifies the requirement for adaptive, site-specific flood mitigation strategies that combine real-time monitoring and sophisticated forecasting models. Future studies should expand the analysis framework to encompass more river basins in contrasting climatic regimes to substantiate these results even more. Additionally, the conjunction of machine learning techniques, recurrent neural networks (RNN) and long short-term memory (LSTM) models, with chaos-induced approaches can improve predictability by modeling intricate temporal interdependencies in precipitation data effectively. Besides, the inclusion of other chaos detection techniques, like the Brock-Dechert-Scheinkman (BDS) test, can enhance statistical verification and reinforce the soundness of chaos identification in hydrological research.

Practically, policymakers and disaster management institutions need to incorporate chaos-based hydrological models in national flood forecasting systems. The use of real-time monitoring systems that integrate chaos-informed predictive models would have the potential for greatly enhancing early warning mechanisms and disaster response measures. Urban environments need to prioritize sustainable drainage practices to curb the risk of flooding, while rural environments need to ensure that natural water retention systems remain intact to effectively manage riverine flow. Interdisciplinary cooperation between hydrologists, climate scientists, and urban planners will be essential to translate these research outcomes into successful policy interventions. By enhancing our knowledge of unstable rainfall patterns, this study opens the door for more adaptive flood management practices, thereby enabling communities to become more resilient to extreme weather events in a time characterized by rising climate variability.

Acknowledgements

We would like to express our gratitude to Jabatan Pengairan dan Saliran (JPS) Malaysia for providing the data used in our study.

References

Ahmed, A. (2023). The Impact of Chaos Theory on Weather Forecasting: Understanding the Complexities of Predicting the Unpredictable. https://www.linkedin.com/pulse/impact-chaos-theory-weather-forecasting-understanding-asghar-ahmed/

Bernardini, D., & Litak, G. (2016). An overview of 0–1 test for chaos. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 38(5), 1433–1450. https://doi.org/10.1007/s40430-015-0453-y

Bhandari, P. (2022). Missing Data | Types, Explanation, & Imputation. https://www.scribbr.co.uk/stats/missing-values/

Biswas, H. R., Hasan, M. M., & Kumar Bala, S. (2018). Chaos Theory and Its Applications in Our Real Life. Barishal University Journal Part, 1(2), 123–140.

Bubeck, P., & Thieken, A. H. (2018). What helps people recover from floods? Insights from a survey among flood-affected residents in Germany. Regional Environmental Change, 18(1), 287–296. https://doi.org/10.1007/s10113-017-1200-y

Domenico, M. De, Ghorbani, M. A., Makarynskyy, O., Makarynska, D., & Asadi, H. (2013). Chaos and reproduction in sea level. Applied Mathematical Modelling, 37(6), 3687–3697. https://doi.org/https://doi.org/10.1016/j.apm.2012.08.018

Drew, C. (2022). 10 Butterfly Effect Examples. https://helpfulprofessor.com/butterfly-effect-examples/

Fauzi, N. U. M., Bakar, M. A.-A. A., Zolkply, N. H., Saleh, S. H. M., Sapini, M. L., & Yusof, N. M. (2022). The application of chaos theory in COVID-19 data analysis. AIP Conference Proceedings, 2472(1), 30011. https://doi.org/10.1063/5.0093272

Forgues, B., & Thietart, R.-A. (2018). Chaos Theory BT - The Palgrave Encyclopedia of Strategic Management (M. Augier & D. J. Teece (eds.); pp. 226–230). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-00772-8_384

Gottwald, G. A., & Melbourne, I. (2016). The 0-1 Test for Chaos: A Review BT - Chaos Detection and Predictability (C. (Haris) Skokos, G. A. Gottwald, & J. Laskar (eds.); pp. 221–247). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-48410-4_7

Hawkins, B. (2011). Sensitivity to Initial Conditions in Chaos. CC BY-NC-SA.

Lu, X., Clements-Croome, D., & Viljanen, M. (2010). Integration of chaos theory and mathematical models in building simulation: Part I: Literature review. Automation in Construction, 19(4), 447–451. https://doi.org/https://doi.org/10.1016/j.autcon.2010.01.002

Peters, E. E. (1996). Fractal Market Analysis: Applying Chaos Theory to Investment and Economics. John Wiley & Sons.

Sapini, M. L., Adam, N. S., Ibrahim, N., Rosmen, N., & Yusof, N. M. (2017). The presence of chaos in rainfall by using 0-1 test and correlation dimension. AIP Conference Proceedings, 1905(1), 50040. https://doi.org/10.1063/1.5012259

Sapini, M. L., Aris, M. N. M., Robangi, A. A., Zulkanain, N. N., & Yusof, N. M. (2019). The existence of chaos in stock market price: An investigation using BDS test and close returns test. AIP Conference Proceedings, 2138(1), 30036. https://doi.org/10.1063/1.5121073

Sivakumar, B. (2000). Chaos theory in hydrology: important issues and interpretations. Journal of Hydrology, 227(1), 1–20. https://doi.org/https://doi.org/10.1016/S0022-1694(99)00186-9

Su, F. (2021). The Chaos Theory and its Application. Journal of Physics: Conference Series, 2012(1), 12118. https://doi.org/10.1088/1742-6596/2012/1/012118

Tella, A., Mustafa, M. R. U., Balogun, A. O., Okolie, C. J., Bello Yamusa, I., & Ibrahim, M. B. (2023). Spatial Prediction Of Flood In Kuala Lumpur City Of Malaysia Using Logistic Regression. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W6-2022, 363–369. https://doi.org/10.5194/isprs-archives-XLVIII-4-W6-2022-363-2023

Application Of Dijkstra's Algorithm To Find The Minimum Travel Time For Public Transport In A City

Nurin Aina Sofhia Binti Zuraidi[1,*], Noor Hidayah Mohd Zaki[1] and Noraimi Azlin Mohd Nordin[2]

*[1*12]Mathematical Sciences Studies, College of Computing Informatics and Mathematics, Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban, Persiaran Seremban Tiga/1, Seremban 3, 70300 Seremban, Negeri Sembilan.*

Corresponding: hidayahzaki@uitm.edu.my

Abstract

Public transportation systems are essential to sustainable urban mobility, particularly in densely populated cities. To achieve this, Dijkstra's Algorithm has been applied to two important objectives: to find the minimum travel time using Dijkstra's Algorithm and to recommend the shortest route for bus transportation services. The transportation network is represented as a graph with nodes representing some key landmarks and edges weighted by travel times and penalties including traffic signals and junctions. Manual calculations in the area of study has been calculated and in order to effectively implement the algorithm, Microsoft Excel and Open Solver were used to provide an expandable, user-friendly platform for computational analysis and visualization. Based on the finding, the shortest path is achieved in 9.30 minutes starting from Node 3 to Node 10. Overall, the results show that Dijkstra's algorithm is a simple yet competent approach in finding the shortest path for bus services within a short time period.

Keywords*:* Transportation, Dijkstra's Algorithm, Microsoft Excel and Open Solver

Introduction

Nowadays, discussing the issue of finding the quickest path in daily life is highly fascinating. As the world's economy grows and the middle class is becoming more prevalent in places like China and India, it is anticipated that the number of vehicles would also rise globally (Indrajaya et al., 2015). Even if the number of transports is growing and the growing middle-class and upper-class population has a good effect, there has not been enough development of vehicle tracking systems, which are crucial components of transportation networks (Indrajaya et al., 2015). Numerous factors, such as economy and safety in private automobiles, public transit systems, commercial vehicles, and so on, will benefit greatly from the deployment of tracking systems on the mode of transportation.

One of the most popular forms of transportation in cities is public transportation. Due to the numerous benefits that public transportation offers its customers, a sizable portion of city residents as well as tourists regularly utilise this urban transportation service. Firstly, compared to other modes of transportation, it is more cost-effective. Furthermore, in some situations such as peak hours, using a taxi may be quicker and less stressful than utilising a personal vehicle. As a result of the large number of individuals who regularly use public transport, multiple route-recommending applications have been developed across various platforms.

One of the common public transportation used are bus services. In Malaysia, this service is considered important and widely used throughout the country. It provides accessible, affordable services for all sections of society, especially for those living in urban and suburban areas. As a flexible form of public transport, buses can travel on various routes and destinations, including areas that may not be accessible by other types of public transport such as trains or MRT.

The process of figuring out the optimal path for a car to take, either before or during travel, to get to its destination is known as path planning. This specific application of the shortest path issue in automobile navigation systems is based on the topological structure of the road network. Bogdan (2015) said, Edsger Dijkstra, a computer scientist, developed Dijkstra's algorithm in 1959 which the idea is to link graph's shortest path problem (SPP). Until the criteria are met, this algorithm will iterate over each point on the graph and determine the shortest path between them. Furthermore, Bogdan (2015) said in practice, the Dijkstra's algorithm determines the shortest routes from the starting point to the other places on a map in addition to the shortest path between the start and finish points. The SPP is often used to minimize travel time and help individuals choose the best route when traveling from one point to another. Dijkstra's algorithm is one method for assisting and resolving the issue of helping individuals choose the shortest route.

One of the common issues arise is road congestion in crowded areas, especially in cities. It is largely driven by the rapid growth of the automobile industry and people's increased purchasing power. The rising number of vehicles, combined with the heavy reliance on private cars, is particularly noticeable during peak hours when people go out for lunch or running daily chores. This surge in vehicle usage often leads to traffic jams, and causing delays for many travelers. One of the solutions to address this issue is by using public transport services, which are highly encouraged by the government, such as buses. However, it is not easy to encourage people to use public transportation. Thus, the use of Dijkstra's algorithm is hoped to give idea to people on how to get to their destination in minimum travel time by using public transportation. This process is represented as a road network, where the edges are connected with nodes.

Based on the process, it is hoped that Dijkstra's algorithm resolves navigational issues. This may also save time and improve decision-making. This study offers a novel framework that enables drivers to efficiently manage their travel based on traffic situation. The findings demonstrate that, given the traffic count, drivers may find both the shortest and alternate paths for a given route. According to experimental data, Makariye (2017) said the algorithm performs better over a range of parameters and responds to the shortest path problem very quickly.

Referring to Bogdan (2015), with applications in timetable information systems, online routes, and navigation systems, algorithm engineering's study of route planning in transportation networks is important. In a weighted graph, the shortest path between nodes is found using Dijkstra's algorithm. The objective is to design a highly efficient parallel application while taking the density and structure of the graph into account. Sequential techniques might be more effective, although the effectiveness of the suggested approach depends on the density and structure of the graph, considering both long and short roads.

According to (Khaing et al., 2018), due to pollution, time wastage, and the usage of hydrocarbon fuels, the traffic issue in the Yangon area lowers the quality of life for its residents. The issue still exists despite an increase in the number of roads. An algorithm known as the shortest path was created to determine the optimal route while taking into account the current

traffic situations and possible detours. The goal of Dijkstra's Algorithm is to identify the shortest path between a source and a destination while solving the SPP in graphs.

According to Ray (2022), vehicle monitoring is crucial for fleet management, mass transit management, public transportation management, and personal safety. Users frequently utilize Dijkstra's algorithm to find the shortest path. Real-time location tracking in a GIS application was created to enhance public transport. Drivers no longer need to drive across town because this system assigns passenger groups to taxis.

As highlighted in Chen (2022), public transport, such as buses and subways, is increasingly important as cities get larger. Using time or distance as a baseline, Dijkstra's algorithm can assist passengers in choosing which station to go between. Assuming that passengers can switch or transfer buses at each station, the algorithm imports roads and stations as nodes and edges. Dijkstra's algorithm is applied to the database processing.

However, the algorithm may be tedious and complicated to be calculated manually if large data is involved. Thus, a software needs to be used to manage the algorithm and to ease in calculations. Microsoft Excel is a versatile and widely accessible tool used for handling complex computations, data analysis, and optimization tasks. Its user-friendly interface and built-in functionalities, such as Visual Basic for Applications (VBA), make it an excellent platform for solving real-world problems. For instance, tools like the VRP Spreadsheet Solver utilize. Erdogan (2017) said, Excel to address routing and scheduling issues by integrating VBA with GIS data, enabling users to calculate travel distances, optimize routes, and visualize solutions dynamically.

On the other hand, Open Solver by Microsoft Excel extends the computational capability of spreadsheet modelling by allowing the use of sophisticated optimization algorithms. Unlike Excel's built-in Solver, Open Solver can handle larger datasets and more complex models, using powerful external solvers like Gurobi, CBC, and CPLEX to compute optimal solutions efficiently.

This paper only focuses on the application of Dijkstra’s algorithm in finding the shortest path of the public bus transportation. The objectives of this paper are firstly to find the minimum travel time from SOGO Kuala Lumpur to Bahtera Sri Kandi Sdn. Bhd. Secondly, to recommend the shortest route for bus transportation services. In the end, minimizing total travel time will contribute towards minimizing the waiting time of the bus passengers.

Methodology

Data Collection

A study was conducted to understand the application of Dijkstra's Algorithm in finding the shortest travel time. This study focused on a simplified network of public transport routes in Kuala Lumpur, using buses as the primary mode of transportation. The analysis was conducted manually at first to allow a deeper understanding of the how the algorithm’s work. This study serves as a foundation for applying the algorithm to a larger dataset, ensuring the methodology is accurate and efficient before expanding to cover more extensive transport networks.

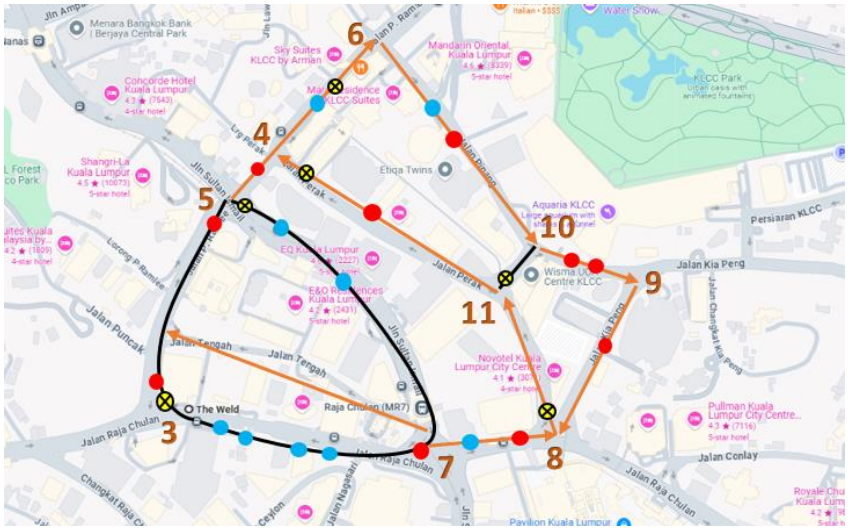


Figure 1: Route network for the area under study

Figure 1 shows the route of respective study. There are nine nodes with different penalties for each route. The starting node is denoted as node 3 and the destination node is node 10. As can be seen in Figure 1, the orange line indicates the one-way direction route. For instance, node 6 to node 10 are indicated by the directional arrows on the edges. Other than that, the routes can be travelled in two-ways direction. Based on this, the algorithm only focusing on feasible routes as set in the objective which is to achieve the minimum travel time. Next, Table 1 listed the description of the nodes. The road names are converted to numbers and on the whole, nine nodes are involved in this study.

Table 1: Description of the nodes involves in area of study

Legend	Description
3	The Weld The Weld
5-4	Jalan P.Ramlee
7	Jalan Raja Chulan
6	Jalan Pinang
8-11	Jalan Perak
9	Jalan Kia Peng
10	Bahtera Sri Kandi Sdn.Bhd

Table 2 describes the legend used in Figure 4 until Figure 9 to show how the Dijkstra's algorithm work to find the minimum travel time by manual calculation. Penalties are presented the same colour as in Figure 1.

Table 2: Description of legends involves in area of study

Legend	Description
3,4,5,6,7,8,9,10,11	Nodes
	Edges two way route
	Edges one way route
	Nodes
	Possible Nodes
	Confirmed Nodes

This study also comes with the assumption that this data:

- Captured on a weekday with typical and normal working hours.
- There are no accidents or construction sites along the route.
- The weather is good.
- The transport is in a good condition.
- The driver is knowledgeable about the roads and will keep to the posted speed limit.
- The data includes the following penalties such as traffic lights, traffic jams, junctions and road bumps.
- The roads taken are main roads only.
- Every visiting sequence has a set destination
- The calculation will begin from The Weld and stop at Bahtera Sri Kandi Sdn. Bhd.

Develop Travel Network

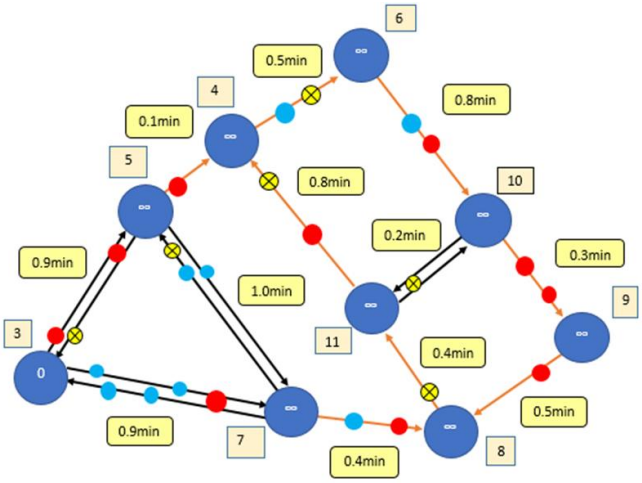


Figure 2: Network for area under study

Based on Figure 2, the network shows a directed graph for one-way journey, with nodes connecting the edges. The weights between the nodes is in minutes. The starting point is node 3 and the destination node is node 10. The nodes and edges have been carefully constructed to represent actual travel limitation, allowing it to compute the shortest journey durations with accuracy and efficiency.

Table 3 presents the travel times between one nodes to another nodes in a network, with each row representing a route from a starting point to an ending point. The table provides five main information which is the starting node, the ending node, travel time (excluding penalties), penalties (additional time caused by factors such as traffic lights, junctions, or bus stops), and the total travel time (a combination of travel time and penalties). For example, traveling from node 3 to node 5 takes 0.9 minutes as the base travel time, but with an additional penalty of 2 minutes, the total travel time becomes 2.9 minutes. This table was calculated manually by inputting values based on collected data, which are calculated manually using the Dijkstra's algorithm to determine the shortest travel time. The calculation of the total travel time with penalties in Table 3 is explained in detail in the next section.

Table 3: Travel Time between nodes in the study

Starting Point	Ending Point	Travel time (minutes)	Penalties (minutes)	Total travel times (minutes)
3	5	0.9	2.0	2.9
3	7	0.9	3.0	3.9
5	7	1.0	3.0	4.0
7	5	1.0	2.5	3.5
5	4	0.1	1.0	1.1
4	6	0.5	1.5	2.0
6	10	0.8	2.5	3.3
7	8	0.4	3.0	3.4
8	11	0.4	2.0	2.4
11	4	0.8	1.5	2.3
11	10	0.2	0.5	0.7

Finding the Shortest Path using Dijkstra's Algorithm

R is a collection of travel nodes initially. R is equal to {A}. Let node A be the source node and the starting point. P(v) is the value of travel time from source node A to node v for all nodes v that are not in R. Each edge in the graph connection from nodes s to nodes v, has the same value of travel time, P(v).

$P(v) = Q(A, v)$, if the nodes v and the source are connected directly.
$P(v) = \infty$, otherwise.

Therefore, the equation is:

$$D(s,v) = Min\left[P(v), P(w) + Q(w,v)\right] \quad (1)$$

where,

- $s =$ source node, where $s = A$
- $D(s, v)$ is the travel distance from node s (source node) to node v.
- $P(v)$ is the value of travel distance from source node 3 to node v for all nodes v that are not in R.
- $P(w)$ is the minimum value of travel distance.
- $Q(w, v)$ is the link cost from node w to node v if two nodes are connected.

In Dijkstra's Algorithm, the shortest path between the starting node and every other node is be determined. Longer travel time are skipped during the modification in order to maintain the shortest path from the starting point. The following steps are provided:

1. All nodes start with "infinite" distance; initialization of the initial node by 0.
2. Label the beginning node's distance as permanent, while all other distances as temporarily.
3. The initial node is active.
4. Compute the temporary distances of all neighboring nodes of the active nodes by adding its distance with the weights of the edges.
5. Update distance and set the node as an antecessor if the measured distance of the node is less than that of the current node. This step is also regarded as an update and is the principal concept of Dijkstra's Algorithm.
6. The minimum temporary distance setting of node as active. Mark the distance as permanent.
7. Redo step four to seven until there are no permanent distance nodes exist, which still have temporary distances for neighbours.

Steps in Dijktra's Algorithm for Finding Minimum Travel Time

In this section, the details of calculating the minimum travel time using the Dijkstra's algorithm is explained deliberately. The manual calculation is hoped to give better understandings in how the algorithm actually works. The procedure will be carried out throughout all the nodes and the algorithm will only stop until all the nodes have been processed. The calculation starts with step 1 as referring to Figure 3.

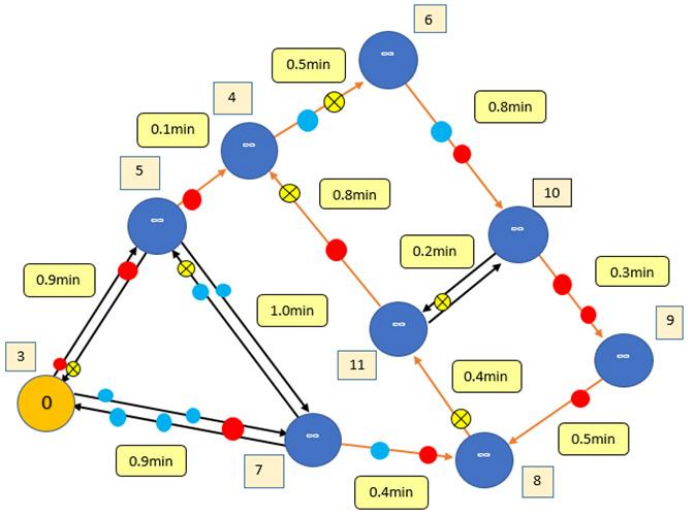


Figure 3: Node 3 is updated in orange

Step 1: Based on Figure 3, orange colour indicates the selected starting node. In this preliminary study, node 3 is set as a starting node. Starting node value is 0. From here, ∞ is assigned to all other nodes. The source node in this study is $R = \{3\}$.

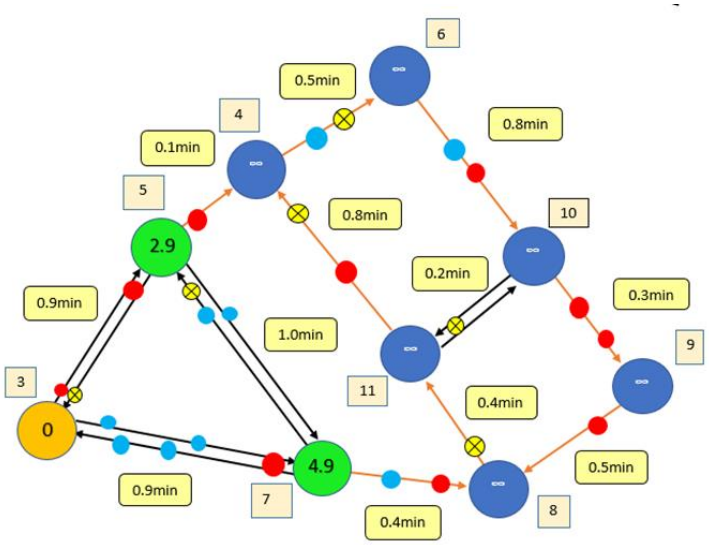


Figure 4: Selecting next possible shortest path from node 3

Step 2: The orange node has been recorded as being visited and will not be rechecked. Based on Figure 4, the next possible shortest path from node 3 in this stage can be either node 5 or node 7, with $P(3) = 2.9\ minutes$ and $P(7) = 4.9\ minutes$, respectively. By comparing the two values, node 5 is likely to be chosen as the current node. The time may be longer if node 3 to node 7 is chosen as the next current node. Thus, $R = \{3, 5\}$ is the current node with a total time of 2.9 minutes with two road penalties. The calculation between node 3 to 5 and node 3 to 7 is explained as follows:

Manual calculation from node 3 to node 5:

As in Figure 2, there are two penalties from node 3 to node 5. Each penalty is equal to additional one-minute travel time. Without the penalties, total travel time is only 0.9 minutes. However, in real life situations, vehicles need to follow strictly the rules abide by the authorities to ensure smooth driving to the destinations.

$\boldsymbol{If, R} = \{\mathbf{3, 5}\}$
$Road\ penalties = 1\ minute + 1\ minute$
$Penalties = 2\ minutes.$
$P(w) = 0.9\ minute + 2\ minutes$

Thus, the current value from node 3 to node 5 is equal to 2.9 minutes. Next, the current travel distance will be updated to $D(3, 5)$. Denotes that the infinity symbol means that there are still nodes to be processed to reach the ending node.
$D(3, 5) = [2.9, (0.9 + 2) + \infty]$
$D(3,5) = 2.9\ minutes$

Hence, current shortest path from node 3 to node 5 is denoted by $R = \{3,5\}$ which equals to 2.9 minutes

Next, comparing the value with manual calculation from node 3 to node 7:

Now in Figure 2, driver may choose either go to node 5 or node 7. Previously, the calculation shows result from node 3 to node 5. Now there are 4 penalties involved in node 3 to node 7. Hence, additional four-minutes will be added up with the current travel time 0.9 minutes.

$$\boldsymbol{If, R} = \{\mathbf{3, 7}\}$$
$$Road\ penalties = 1\ minute + 1\ minute + 1\ minute + 1\ minute$$
$$Penalties = 4\ minutes.$$
$$P(w) = 0.9\ minute + 4\ minutes$$

Thus, the total travel time is now 4.9 minutes. As can be seen, it takes two minutes longer compared to the other node selection. Therefore, the current travel distance is updated to $D(3,7)$ with the total travel time of 4.9 minutes.

$$D(3,7) = [4.9, (0.9 + 4) + \infty]$$
$$D(3,7) = 4.9\ minutes$$

Step 3: As a conclusion, the next selected shortest path is updated to node 5. Now node 5 is colored orange indicates that the node has been chosen as current node.

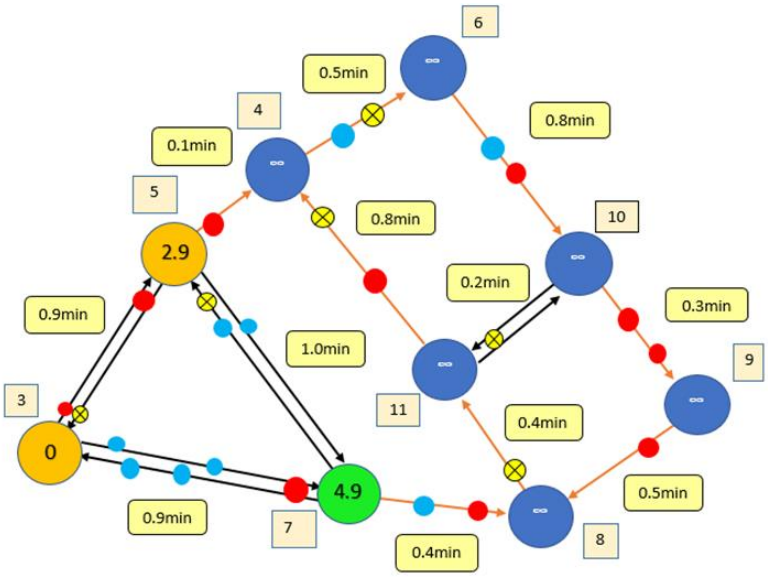


Figure 5: Node 5 updated as selected node (orange colour)

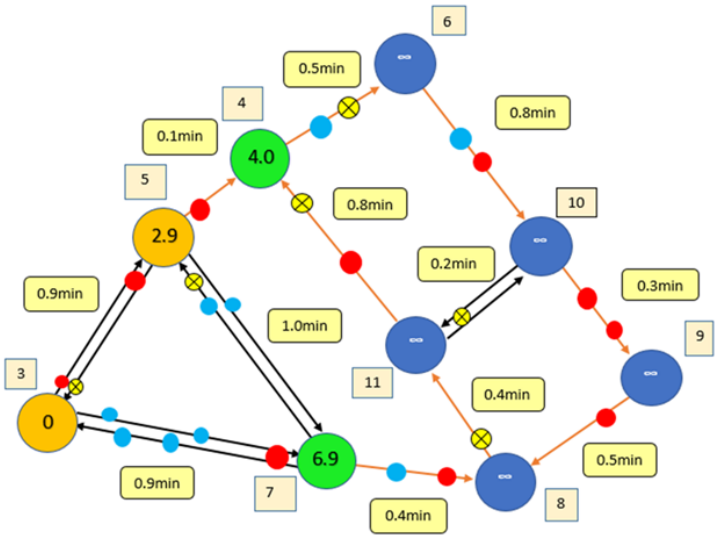


Figure 6: Node 5 to possible next node for travel time

Step 4: The same step is repeated again to find the next possible shortest path from node 5. Figure 6 shows that node 4 and node 7 is the next possible node.

In this situation, there is one road penalty from node 5 to node 4 and three penalties from node 5 to node 7. Now, the travel nodes R is updated to either $R = \{3,5,4\}$ and $R = \{3,5,7\}$. The calculation is explained as follows:

Manual calculation from node 5 to node 4.

$$\boldsymbol{If, R = \{3,5,4\}}$$
$$Road\ penalties = 1\ minutes,$$
$$P(w) = 0.1\ minute + 2.9\ minutes + 1\ minute$$

The calculation of $P(w)$ represent that 0.1 minute is the direct travel from node 5 to node 4, 2.9 minutes is the current travel time value and 1 minute is the penalty involved. Hence, the total travel time is now updated to $D\{3,5,4\}$.

$$D\{3,5,4\} = Min[4.0, (0.1 + 1) + 2.9]$$
$$D\{3,5,4\} = 4.0\ minutes$$

On the other hand, the manual calculation from node 5 to node 7 is as follows:

$$\boldsymbol{If, R = \{3,5,7\}}$$
$$Road\ penalties = 1minute + 1minute + 1\ minute,$$
$$P(w) = 2.9\ minute + 1\ minute + 1minute + 1\ minute + 1\ minute$$

The calculation of $P(w)$ now represent that 1 minute is the direct travel from node 5 to node 7, 2.9 minutes is the current travel time value and 3 minutes are the penalty involved. Hence, the total travel time is now updated to $D\{3,5,7\}$.

$$D\{3,5,7\} = Min[6.9, (1.0 + 3) + 2.9]$$
$$D\{3,5,7\} = 6.9\ minutes$$

ELSE,

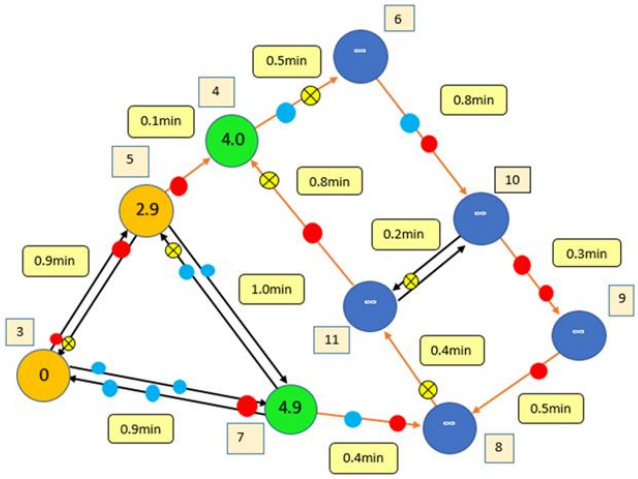


Figure 7: Node 3 to node 7 for another possible shortest node

Step 5: Since Figure 7 is still not selected as the shortest path, it can still be considered as the next possible routes. Earlier in step 3, the travel time from node 3 to node 7 is 4.9 minutes which is shorter than $D\{3,5,7\}$. Due to this, the current shortest path is updated to node 3 to node 7 which is 4.9 minutes. It is still possible since node 3 to node 7 is not selected yet and it can still be processed as possible minimum travel time. There are three choices of next shortest path that are $D\{3,5,4\} = 4\ minutes$, $D\{3,5,7\} = 6.9\ minutes$ or $D\{3,7\} = 4.9\ , minutes$. Based on the choices, $D\{3,5,4\}$ has the shortest value and thus, $D\{3,5,4\}$ is updated to the latest shortest path and is updated as orange colour as in Figure 8.

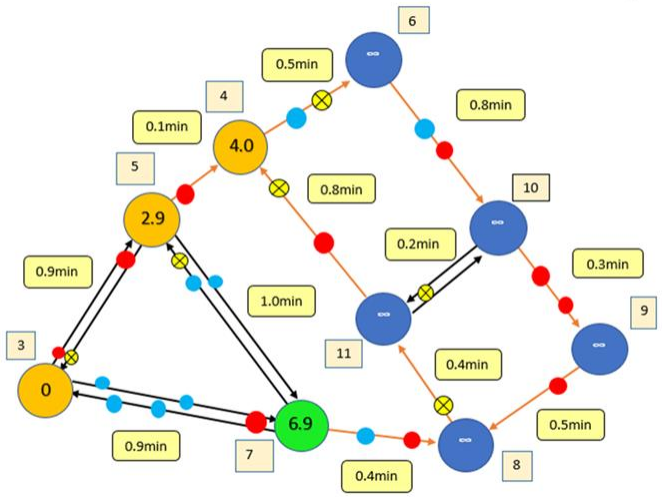


Figure 8: Node 4 is updated as the next shortest path

Step 6: Node 4 is currently selected as shortest path as in Figure 8. Therefore $R = \{3,5,4\}$ with the total of 4 minutes.

Same procedure will be carried out throughout findings shortest path problem until all the nodes has been processed. The calculation stops once the minimum travel time is obtained. After all the shortest path has been updated, the colours are changed to orange. Figure 9 shows the completed network with all the nodes has been processed.

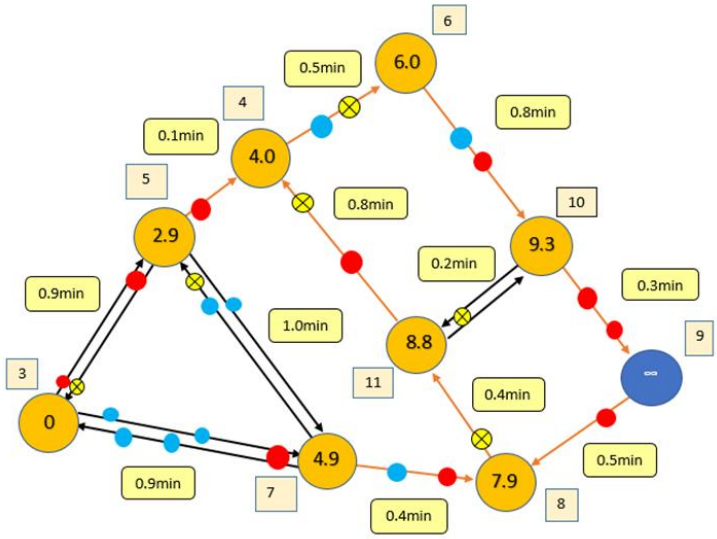


Figure 9: All nodes have been processed and completed

Finally, based on the manual calculation, it is found out that the shortest path from node 3 to node 10 is achieved via path 3 → 5 →4 → 6 → 10. It can also be represented as in the R terms where $R = \{3,5,4,6,10\}$ with 9.3 minutes including all the road penalties. However, doing manual calculations can be tedious and time consuming. It also may increase the error in calculations. Due to this, the help of software is useful since people needs to have fast result in just a matter of time. Thus, in this study, the Microsoft Excel solver is used to find the minimum travel time.

Microsoft Excel Solver for Finding Minimum Travel Time

In this study, Microsoft Excel Solver was utilized as the primary tool for solving the optimization model related to route planning. Solver was chosen due to its capability to handle complex optimization problems efficiently and its integration with Excel, which allowed seamless data management and computation. By using Solver, the optimal travel routes and their corresponding times were determined, forming the foundation for further analysis and decision-making in this research. This preliminary study aimed to validate the feasibility of the proposed approach and ensure accurate results for subsequent stages of the study.

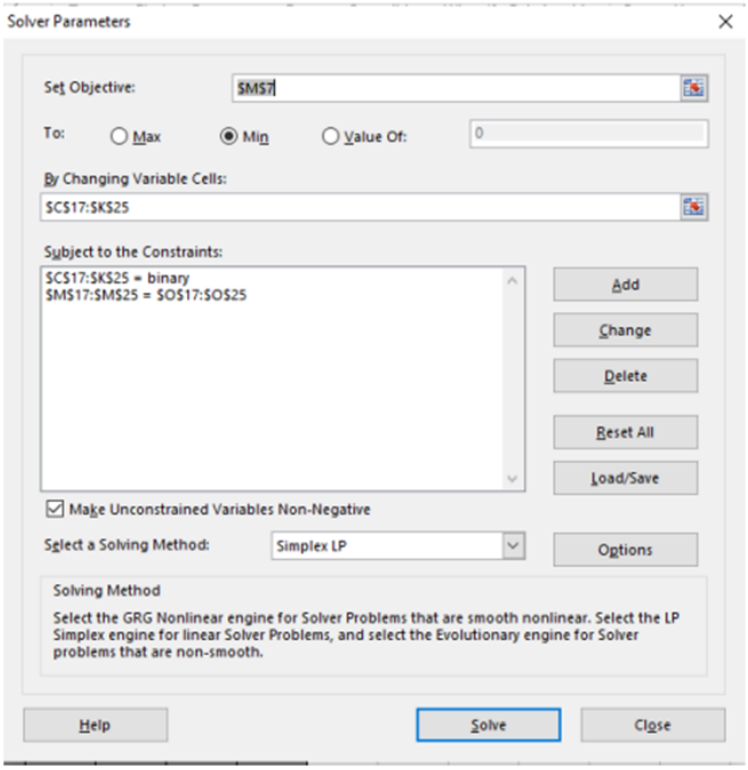


Figure 10: Microsoft Excel solver to find the shortest route with minimum travel time

After setting up the tables in Excel (Table 4), use the Excel Solver to find the shortest route with the least travel distance. Click the Solver button (blue line) to open the Solver dialog box, as

shown in Figure 10. Set the objective to "Min" to minimize travel distance, and select the nodes column as the variable cells in the Excel which is the empty column and row before the result were filled as in Table 5.

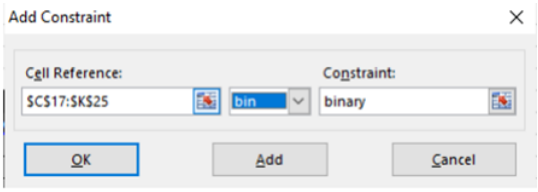


Figure 11: Solver box to add constraints

Next, add constraints in the solver box. Figure 11 shows the first constraint, where the nodes column is set as all different so that there are no repeated nodes. Constraint two was set the starting point and ending point must be equal to one and relate with the $SUMIF$ function. The solving methods are set to Simplex LP and can be solved by clicking the Solve button. The Total Travel Time will be automatically appeared follows the data that has been filled.

The data obtained from area of study has been applied in Microsoft Excel and based on the steps shown. The results obtained are the same as manual calculation in preliminary study which is from $3-5-4-6-10$ with a total travel time of 9.3 minutes. This can be seen in Table 4.

Table 4: Travel time from one to one connected node

		TO								
		3	5	4	6	7	8	9	10	11
FROM	3	100	2.9	100	100	3.9	100	100	100	100
	5	2.4	100	1.1	100	4	100	100	100	100
	4	100	100	100	2	100	100	100	100	100
	6	100	100	100	100	100	100	100	3.3	100
	7	4.9	3.5	100	100	100	3.4	100	100	100
	8	100	100	100	100	100	100	100	100	2.4
	9	100	100	100	100	100	2.5	100	100	100
	10	100	100	100	100	100	100	2.3	100	0.7
	11	100	100	2.3	100	100	100	100	0.7	100

Table 4 shows the setup of calculation to find minimum travel time for preliminary study. The column besides the column "FROM" represent as nodes, below the words "TO" also represent as nodes which it will read from left to right. The number of 100 means there is no route between the two nodes, such as from node 3 to node 4, 100 was filled like shows in the Table 4. If the route from one node to other nodes is exist the value of travel time will fill in the table such as from node 3 to node 5 there is route between the two nodes which is 2.9 minutes.

Once the settings are done, the calculation can be obtained by just choosing the starting and ending point. The result based on the Microsoft Excel is discussed and compared with the manual calculation under the Result and Discussion section.

Result and Discussion

All the data pertaining is focusing on the bus transportation in Kuala Lumpur. From there, a complete road network is developed. Hence, once the starting node and the destination node is identified, a detailed road network that shows complete travel times is developed as in Figure 2. In this case, SOGO Kuala Lumpur is the starting node (node 3) and Bahtera Sri Kandi Sdn. Bhd. (node 10) is the ending node whereby the other nodes is connected by edges.

Based on the route network, there are many possibilities to reach node 10. The main concern is whether the route chosen is the fastest. Since in this case, people or the passanger will choose to ride in a less crowded and less congested bus. Other than that, passengers aim at arriving earlier at the designated place. Therefore, shortest path is vital. Thus, the Dijkstra's algorithm is applied in determining the minimum travel time. The results of this algorithm is explained in details in this section.

Table 5: Result of the minimum travel time from node 3 to node 10

		TO												
		3	5	4	6	7	8	9	10	11	TOTAL OUT	OUT - IN	=	
FROM	3	0	1	0	0	0	0	0	0	0	1	1	=	1
	5	0	0	1	0	0	0	0	0	0	1	0	=	0
	4	0	0	0	1	0	0	0	0	0	1	0	=	0
	6	0	0	0	0	0	0	0	1	0	1	0	=	0
	7	0	0	0	0	0	0	0	0	0	0	0	=	0
	8	0	0	0	0	0	0	0	0	0	0	0	=	0
	9	0	0	0	0	0	0	0	0	0	0	0	=	0
	10	0	0	0	0	0	0	0	0	0	0	-1	=	-1
	11	0	0	0	0	0	0	0	0	0	0	0	=	0
	TOTAL IN	0	1	1	1	0	0	0	1	0				

Table 5 shows the result after run the open solver. The last column shows that there is $1, -1$ and 0.1 are represent as the first node and -1 represent as last node, other nodes was filled with 0. Shows that in the column "Total OUT", the selected node chosen after run the Excel Solver The "Total OUT" column in shortest route problems using Microsoft Excel Solver shows the number of outgoing connections from each node. In a directed graph, where each path can only be used once, a value of "1" indicates an active connection selected by Solver. Since paths are non-repetitive, Solver ensures each connection is unique and forming a valid route. Constraints like "OUT - IN = 0" for intermediate nodes and "OUT - IN = ±1" for start and end nodes maintain route integrity. For example, if Node 3 has a Total OUT of 1, it means Solver selected one outgoing path from Node 3 as part of the shortest route.Number 1 (yellow) shows the confirmed shortest route from node 3 to node 10 other than another route.

TOTAL TRAVEL TIME
9.3

Figure 12: Total Travel Time for preliminary study

Objective function in the excel solver in Figure 12 refers to the solution which is to find the minimum travel time. The result using Microsoft excel shows the total value 9.3 minutes as the from node 3 to node 10.

Based on the comparison between manual calculation and execution using Microsoft Excel software in the previous section, it shows that the minimum travel time from SOGO Kuala Lumpur to Bahtera Sri Kandi Sdn. Bhd is achieved via path $3 \rightarrow 5 \rightarrow 4 \rightarrow 6 \rightarrow 10$, with the total distance of 9.3 minutes. The result shows that the manual calculation is correct since both procedures obtained the same result. This proves that the Dijkstra's algorithm is stable and easy to compute due to the straightforward process.

Conclusion

The study has thus successfully applied Dijkstra's Algorithm in finding the minimum travel time from SOGO Kuala Lumpur to Bahtera Sri Kandi Sdn. Bhd. The algorithm estimated realistic travel times once real factors like traffic lights, junction penalties, and delays at bus stops were added. This approach arrived at the best routes with the minimum time taken, hence assisting in better planning for optimization of bus routes. The study also succeeded in recommending the ideal routes of buses with a minimum of overall travel time in Malaysia, taking into consideration essential factors affecting urban mobility. These recommended routes minimize the time spent in traveling at traffic lights, junctions, and stops.

References

Bogdan, P. (2015). Dijkstra Algorithm in Parallel- case study. Dijkstra Algorithm in Parallel. https://doi.org/10.1109/carpathiancc.2015.7145044

Chen, R. (2022). Dijkstra's Shortest Path algorithm and its application on bus routing. Advances in Economics, Business and Management Research/Advances in Economics, Business and Management Research. https://doi.org/10.2991/aebmr.k.220502.058

Erdogan, G. (2017). An open-source Spreadsheet Solver for Vehicle Routing Problems. Computers & Operations Research, 84, 62–72. https://doi.org/10.1016/j.cor.2017.02.022

Indrajaya, M. A., Affandi, A., & Pratomo, I. (2015). Design of Geographic Information System for tracking and routing using Dijkstra Algorithm for public transportation. 2015 1st International Conference on Wireless and Telematics (ICWT). https://doi.org/10.1109/icwt.2015.7449240

Khaing, O., Wai, H. H., & Myat, E. E. (2018). Using Dijkstra's algorithm for public transportation system in Yangon based on GIS. International Journal of Science and Engineering Applications, 7(11), 442–447. https://doi.org/10.7753/ijsea0711.1008

N. Makariye, "Towards shortest path computation using Dijkstra algorithm," 2017 International Conference on IoT and Application (ICIOT), Nagapattinam, India, 2017, pp. 1-3, https://doi.org/10.1109/ICIOTA.2017.8073641

Mason, A., & Dunning, I. (2010). OpenSolver: Open source optimisation for Excel. Proceedings of the Annual Conference of the Operations Research Society of New Zealand.

Ray, A. (2022). Analysis and design of public transport Route planner: Dijkstras Algorithm. International Journal for Research in Applied Science and Engineering Technology, 10(6), 4571–4575. https://doi.org/10.22214/ijraset.2022.45018

Shortest Route Optimization for Emergency Cases using Integer Linear Programming

Nurul Liyana Abdul Aziz*, Aisyah Fariha Norzaki
Fakulti Sains Komputer dan Matematik,
Universiti Teknologi MARA Negeri Sembilan, Kampus Seremban
*liyana511@uitm.edu.my

Abstract

Emergency cases require efficient route planning to minimize response times and ensure timely access to critical care. This study presents an optimization model for determining the shortest route in emergency cases using Integer Linear Programming (ILP). The model aims to efficiently allocate resources such as ambulances or private vehicles by optimizing their routes while considering the road networks. This study considers six cities and six hospitals in Seremban, with key nodes representing cities, hospitals, and major intersections, whiles edges representing the connecting roads. 36 case studies are conducted using real-world road network data, and the solution is obtained using QM for Windows and Excel Solver. Results demonstrate that the distance from each city to Hospital Tuanku Ja'afar which is the government hospital is in the range of 4.8km to 8.2km. However, for alternatives, the results show that the nearest hospital to Seremban is CMH Specialist Hospital. For Oakland and Seremban 2, the closest hospital is Columbia Asia Hospital. Mawar Medical Centre serves as the nearest hospital for Seremban 3 and Taman Rahang. Meanwhile, residents of Ampangan have CMH Specialist Hospital and SALAM Senawang Specialist Hospital as their nearest medical facilities. This study contributes to the development of intelligent decision-support systems for emergency management, enhancing rapid response capabilities and overall public safety.

Keywords*:* Shortest route problem, Integer Linear Programming, QM for Windows, Excel Solver

Introduction

In an emergency, every second matters. To resolve the emergency cases issue, the criteria by which drivers select the routes between their starting point and destination must be specified. In most cases, every driver may wish to minimize the route to the destination. The shortest route problem can simply be defined as finding the most efficient route between two points in a graph where this efficiency is measured by many variables such as distance, time, and cost.

This study deals with the problem of finding shortest routes in traversing some locations in Seremban for emergency cases. The shortest route will directly reduce the emergency vehicle's response time. For example, a quicker response time corresponds with a higher patient survival rate. Therefore, the response time of the rescue vehicle should be optimized by determining the shortest route to the emergency center. Integer Linear Programming (ILP) has emerged as a powerful mathematical tool to address the complexities involved in routing emergency vehicles under various constraints.

Previously, Li et al. [1] developed an integrated emergency vehicle scheduling model tailored for urban emergency rescue scenarios within constrained time frames. This model incorporates real-time traffic data and prioritizes critical cases to ensure rapid response. The study demonstrates that integrating live traffic updates can significantly improve response times. However, the reliance on real-time data introduces computational complexities, which may limit its feasibility in resource-constrained emergency response systems.

Besides that, Wang et al. [2] addressed the routing of emergency vehicles on highways by developing a mixed-integer linear programming model that considers factors such as traffic congestion, incident severity, and vehicle availability. The model aims to minimize total response time, maximize the success rate of emergency responses, and optimize the utilization of emergency vehicles. Tested on real-world datasets, the model demonstrated effectiveness in improving emergency vehicle routing on highways. While the study provides

valuable insights, its focus on highway scenarios may limit its applicability to urban environments with more complex road networks and variable traffic patterns.

Similarly, García et al. [3] developed a mixed-integer linear programming model to optimize the relocation of Emergency Medical Vehicles (EMV) in Valencia, Spain. The study aimed to minimize the population not covered within the stipulated maximum response times of 12 and 15 minutes for advanced life support and basic life support units, respectively. By proposing a new distribution of EMV bases, the model reduced the coverage deficit by more than half compared to the existing arrangement. This research underscores the significance of strategic vehicle placement in enhancing emergency response efficiency. However, the study primarily focuses on static relocation strategies and does not account for real-time traffic conditions or dynamic demand fluctuations, which are critical in actual emergency scenarios.

Other than that, Nodoust et al. [4] explored the vehicle routing problem for humanitarian relief distribution under hybrid uncertainty, combining both probabilistic and possibilistic approaches to model demand uncertainty. They developed a robust scenario-based possibilistic-stochastic programming model to distribute first aid relief items in post-disaster scenarios where routes are subject to disruption. Applied to a real earthquake case study, the model demonstrated a higher demand satisfaction level compared to traditional stochastic programming models. This study highlights the importance of considering hybrid uncertainty in emergency logistics. However, the complexity of the model may pose challenges for practical implementation in time-sensitive disaster response situations.

The application of linear programming in optimizing emergency response routes has shown promising results in various contexts. By addressing existing gaps and building upon previous research, future studies can develop more robust and adaptable models that enhance the efficiency and effectiveness of emergency services across diverse scenarios.

Network Problem Definition

This study starts with identifying and defining the nodes and pathways that form the network for the emergency services. The nodes in this study are the selected cities and hospitals located in Seremban.



Figure 1: Map of selected area

Figure 1 illustrates the study area in Seremban, highlighting the orange nodes representing cities and blue nodes representing hospitals. These nodes serve as key points in the optimization model for determining the shortest routes for emergency services. This study focuses on six cities and six hospitals as in Table 1 and there are 36 emergency cases are considered.

Table 1: Cities and hospitals

Cities		Hospitals	
C1	Seremban	H1	CMH Specialist Hospital
C2	Oakland	H2	KPJ Seremban Specialist Hospital
C3	Seremban 2	H3	Columbia Asia Hospital
C4	Seremban 3	H4	Mawar Medical Centre
C5	Taman Rahang	H5	Hospital Tuanku Ja'afar
C6	Ampangan	H6	SALAM Senawang Specialist Hospital

Figure 2 shows the road network for Case 1 spans from C1 to H1. The key intersections marked as junctions J1 through J11. The network for Case 1 consists of 13 nodes are constructed with the goal of optimizing the route for transportation.

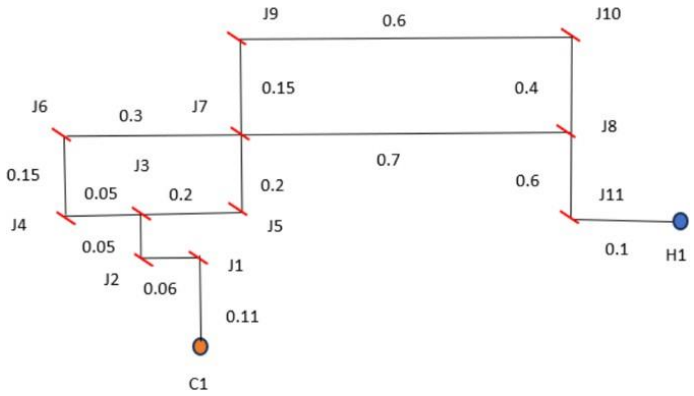


Figure 2: Road network from C1 to H1

ILP Model Formulation for Case 1

To formulate the ILP model for the optimization of the shortest route for emergency services, consider the definition of decision variables, an objective function, and constraints.

Parameters

c_j the distance (km)
$A(i)$ the set of the arc that leave node i
$B(i)$ the set of the arc that enter node i
x_j the flow on arc j

Decision variables

$$\textbf{Let } x_j = \begin{cases} \mathbf{1}, & \textbf{if route } j \textbf{ is chosen} \\ \mathbf{0}, & \textbf{otherwise} \end{cases}$$

x_1	C1 to J1	x_8	J6 to J7
x_2	J1 to J2	x_9	J7 to J8
x_3	J2 to J3	x_{10}	J7 to J9
x_4	J3 to J4	x_{11}	J9 to J10
x_5	J3 to J5	x_{12}	J10 to J8
x_6	J4 to J6	x_{13}	J8 to J11
x_7	J5 to J7	x_{14}	J11 to H1

Objective function

The objective is to minimize the total travel distance

$$\text{Min} \sum_j c_j x_j \tag{1}$$

$$\text{Min Z} = 0.11x_1 + 0.06x_2 + 0.05x_3 + 0.05x_4 + 0.2x_5 + 0.15x_6 + 0.2x_7 + 0.3x_8 + 0.7x_9 + 0.15\,x_{10} + 0.6x_{11} + 0.4x_{12} + 0.6x_{13} + 0.1x_{14}$$

Constraints

Below are the flow conservation constraints for each node in Case 1 network.

Node 1:

$$\sum_{j \in A(i)} x_j = 1 \tag{2}$$

$$x_1 = 1$$

Node 2 to Node 12:

$$\sum_{j \in A(i)} x_j - \sum_{j \in B(i)} x_j = 0 \tag{3}$$

Node 2: $x_1 = x_2$

Node 3: $x_2 = x_3$

Node 4: $x_3 = x_4 + x_5$

Node 5: $x_4 = x_6$

Node 6: $x_5 = x_7$

Node 7: $x_6 = x_8$

Node 8: $x_7 + x_8 = x_9 + x_{10}$

Node 9: $x_9 + x_{12} = x_{13}$

Node 10: $x_{10} = x_{11}$

Node 11: $x_{11} = x_{12}$

Node 12: $x_{13} = x_{14}$

Node 13:

$$-\sum_{j \in B(i)} x_j = -1 \tag{4}$$

$$x_{14} = -1$$

Results and Discussion

QM for Windows - [Data] Results

C1 to H1 solution

Total distance = 2.02	Start node	End node	Distance	Cumulative Distance
J 1	1	2	.11	.11
J 2	2	3	.06	.17
J 3	3	4	.05	.22
J 5	4	6	.2	.42
J 7	6	8	.2	.62
J 8	8	9	.7	1.32
J 11	9	12	.6	1.92
H 1	12	13	.1	2.02

Figure 3: Result of QM for Windows from C1 to H1

Figure 3 shows the result of QM for Windows from C1 to H1 by identifying the optimal junctions. The identified path includes the following junctions: C1 → J1 → J2 → J3 → J5 → J7 → J8 → J11 → H1 with the total distance of 2.02 km.

To verify the results obtained using QM for Windows, Excel Solver was used. Figure 4 presents the data layout in Excel. The decision variables are in the changing cells A2:A15, while cell D16 serves as the objective cell. Cell D16 contains the summation of the decision variables (0 or 1) multiplied by their corresponding distances. Based on Figure 4, the total distance from C1 to H1 is 2.02 km as the result obtained from QM for Windows.

	A	B	C	D
1	Select	From	To	Distance
2	1	C1	J1	0.11
3	1	J1	J2	0.06
4	1	J2	J3	0.05
5	0	J3	J4	0.05
6	1	J3	J5	0.2
7	0	J4	J6	0.15
8	1	J5	J7	0.2
9	0	J6	J7	0.3
10	1	J7	J8	0.7
11	0	J7	J9	0.15
12	0	J9	J10	0.6
13	0	J10	J8	0.4
14	1	J8	J11	0.6
15	1	J11	H1	0.1
16		TOTAL		2.02

Figure 4: Data layout in Excel

Figure 5 illustrates the constraints layout in Excel for each node from C1 to H1. As for out flow of Node 1, cell G2 represents the formula "=SUMIF($B$2:$B$15,F2,$A$2:$A$12)" while for in flow of Node 1, cell H2 represents the formula "=SUMIF($C$2:$C$15,F2,$A$2:$A$15)". The net flow of a node is determined by subtracting its inflow from its outflow.

The results obtained in Column K in Figure 5 satisfy the conditions and match the solution obtained from QM for Windows. This consistency validates the optimization model and its adherence to the defined constraints, ensuring an accurate and feasible solution.

	F	G	H	I	J	K
1	Node	Outflow	Inflow	Netflow		
2	C1	1	0	1	=	1
3	J1	1	1	0	=	0
4	J2	1	1	0	=	0
5	J3	1	1	0	=	0
6	J4	0	0	0	=	0
7	J5	1	1	0	=	0
8	J6	0	0	0	=	0
9	J7	1	1	0	=	0
10	J8	1	1	0	=	0
11	J9	0	0	0	=	0
12	J10	0	0	0	=	0
13	J11	1	1	0	=	0
14	H1	0	1	-1	=	-1

Figure 5: Constraints layout in Excel

Table 2 summarizes the resulted distances from six selected cities to six selected hospitals. H5, a government hospital, is considered the primary choice for each city due to its role in providing affordable and accessible healthcare. The distance from each city to H5 ranges from 4.8 km to 8.2 km. However, as an alternative, residents of C1 can be transported to H1, which is 2.02 km away. For residents of C2 and C3, the nearest alternative hospital is H3, located 0.45 km and 4.20 km away, respectively. Meanwhile, the alternative hospital for residents of C4 and C5 is H4, with distances of 6.20 km and 4.10 km, respectively. For residents of C6, two private hospital options are available: H1 and H6, with a minimum distance of 7.2 km.

Table 2: Shortest distance for 36 emergency cases

From	To	Shortest distance (*km*)	From	To	Shortest distance (*km*)
C1	**H1**	**2.02**	C4	H1	10.20
	H2	3.60		H2	7.90
	H3	4.80		H3	7.70
	H4	5.85		**H4**	**6.20**
	H5	6.55		H5	6.80
	H6	13.35		H6	10.70
C2	H1	8.40	C5	H1	5.20
	H2	5.10		H2	5.50
	H3	**0.45**		H3	6.10
	H4	4.80		**H4**	**4.10**
	H5	5.40		H5	4.80
	H6	12.8		H6	9.30
C3	H1	8.30	C6	**H1**	**7.20**
	H2	5.10		H2	8.80
	H3	**4.20**		H3	9.90
	H4	6.20		H4	7.50
	H5	7.30		H5	8.20
	H6	13.80		**H6**	**7.20**

Conclusion

In conclusion, this study successfully demonstrated the effectiveness of Integer Linear Programming in optimizing emergency service routes within six cities and six hospitals in Seremban. The developed model minimized travel distances, was validated through multiple tools (QM for windows and Excel Solver), and proved reliable in enhancing emergency response efficiency. These findings highlight the practical value of

mathematical modeling in improving healthcare logistics and urban planning, while also providing a strong basis for future studies on emergency route optimization.

References

[1] Li, X., Wang, X., Xu, J., Chen, W., & Lyu, B. (2024). Integrated emergency vehicle scheduling model in urban emergency rescue within a time-limited period. *Transportation research record*, *2678*(7), 802-818.

[2] Wang, Y., Tian, J., Li, J., Cao, Y., & Zhang, J. (2023). A mixed-integer linear optimization problem for highway emergency vehicle routing. In *Eighth International Conference on Electromechanical Control Technology and Transportation (ICECTT 2023)*, 12790, 1370-1377.

[3] García, M. Á. V., Juliá, M. F. V., Regalado, E. V., & Krylova, Y. K. (2022). A mixed integer linear program for a real relocation problem of emergency medical vehicles in the province of Valencia. *Journal of Applied Research in Technology & Engineering*, *3*(2), 85-92.

[4] Nodoust, S., Pishvaee, M. S., & Seyedhosseini, S. M. (2023). Vehicle routing problem for humanitarian relief distribution under hybrid uncertainty. *Kybernetes*, *52*(4), 1503-1527.

FINANCIAL PERFORMANCE EVALUATION OF COMPANIES USING TOPSIS APPROACH: A CASE STUDY OF PROPERTIES SECTOR IN MALAYSIA

Ariesa Sofea Mohd Yusof[1], Farah Waheeda Azhar[2,*]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Negeri Sembilan Branch, Seremban Campus, Seremban 3, 70300 Seremban, Negeri Sembilan, Malaysia*
[2]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Perak Branch, Tapah Campus, 35400 Tapah, Perak, Malaysia*
*farah154@uitm.edu.my

Abstract

The emergence of the property sector in Malaysia has been driven by the increasing of demand from consumers, reflecting the nation's economic growth. Therefore, this study aims to assess the financial performance and rank the financial performance of the listed property companies in Malaysia using TOPSIS model. The data of this study consists of Sime Darby Construction Berhad, S P Setia, Glomac Berhad, SkyWorld Development Berhad, LBS Bina Group Berhad are listed property companies in Malaysian stock market. The results of this study show that SkyWorld Development Berhad achieves the first ranking, followed by Sime Darby Property Berhad, Glomac Berhad, LBS Bina Group Berhad and S P Setia Berhad within the study period of the year 2023. This study is significant because it helps to evaluate, compare and rank the financial performance of the listed property companies in Malaysia with the proposed conceptual framework based on TOPSIS model.

Keywords: TOPSIS Method, Property sector, Financial performance indicators, Decision making, Investment

Introduction

The property sector in Malaysia is an important part of Malaysia that has contributed immensely to the economic growth employment opportunities and investment opportunities. As this sector continues to evolve amidst changing market dynamics and economic conditions, the evaluation of its financial performance becomes increasingly critical.

Based on study by [11], the liberalization of the Malaysian property sector has sparked a lot of interest in foreign investors. However, the influence of the COVID-19 pandemic affects to the property developers not only in Malaysia but also Indonesia has emerged a rising study area [5] . Additionally, a study from [7] also has explores the various shifts in the property market and the dynamics of the study subject emerged from the pandemic. Thus, embracing constant study and analysis of the property market also is instrumental in the fulfilment of the objectives of the current and potential stakeholders and investors. Hence, in the context of the financial performance evaluation, the most ideal method for this evaluation is the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) approach. TOPSIS is a decision-tool which aims to solve multi-criteria decision-making problem [2]. The application of TOPSIS for finding the financial performance in different sectors has been applied in various case studies, such as the selection of logistics providers, [1], supplier selection [10] and evaluation of financial institutions of commercial banks in India [3] that showcases the versatility and effectiveness of this method in diverse settings.

One of the main advantages of TOPSIS over other methods, such as Analytic Hierarchy Process

(AHP) or Data Envelopment Analysis (DEA), is its ability to provide a clear ranking of alternatives while considering both the shortest distance from the ideal solution and the farthest distance from the negative ideal solution. Unlike AHP, which can be subjective in assigning weights, or DEA, which focuses primarily on efficiency measurement, TOPSIS integrates multiple criteria and produces results that are straightforward and easy to interpret. Moreover, TOPSIS can handle complex decision-making environments, including group decision-making, multi-criteria evaluation, and human resource selection systems[4]. In a study of [8], the author compares the performance of manufacturing companies in the U.S. market the authors employed the TOPSIS technique using 11 financial ratios. Other than that, it also makes it easier to rank companies according to their financial performances that is enable comparisons between companies and create a structured way of evaluating performances [9]. Therefore, it is useful for property companies to calculate financial performance as they strategies on how best to cut costs and improve their performance in the face of rising competition.

Data and Methodology

The financial performance of the five listed property companies in 2023 was assessed using ten financial ratios covering liquidity, leverage, profitability, market standing, and efficiency. Beneficial indicators included CR, ROE, ROA, Net Profit Margin, P/E Ratio, EPS, Dividend Yield, and ART, while Debt-to-Equity and Debt-to-Assets were treated as cost criteria to be minimized. This classification allowed the TOPSIS method to rank the companies accurately based on their overall financial strength.

Table 1: Companies used in the study

Property Company	**Code**
Sime Darby Property Berhad	C1
LBS Bina Group Berhad	C2
S P Setia Berhad	C3
Glomac Berhad	C4
SkyWorld Development Berhad	C5

The formula for the listed financial performance indicators used in the study are shown in from (1) to (10):

$$\text{Return on equity} = \frac{\text{Net profit}}{\text{Total shareholders equity}} \times 100\% \tag{1}$$

$$\text{Return on assets} = \frac{\text{Net profit}}{\text{Total assets}} \tag{2}$$

$$\text{Profit margin} = \frac{\text{Net profit}}{\text{Sales}} \times 100\% \tag{3}$$

$$\text{Current ratio (CR)} = \frac{\text{Current assets}}{\text{Current liabilities}} \times 100\% \tag{4}$$

$$\text{Debt to equity ratio} = \frac{\text{Total liabalities}}{\text{Total shareholders equity}} \times 100\% \tag{5}$$

$$\text{Debt to assets ratio} = \frac{\text{Total debts}}{\text{Total assets}} \times 100 \tag{6}$$

$$\text{Price earnings ratio} = \frac{\text{Market price per share}}{\text{Earnings per share}} \tag{7}$$

$$\text{Earnings per share} = \frac{\text{Net profit}}{\text{Number of shares}} \times 100\% \tag{8}$$

$$\text{Dividend yield} = \frac{\text{Dividend per share}}{\text{Market price per share}} \times 100\% \tag{9}$$

$$\text{Account receivables turnover (ART)} = \frac{\text{Net credit sales}}{\text{Average accounts receivable}} \tag{10}$$

TOPSIS model aims to determine the alternative which is the closes to the best ideal solution and farthest from the worst ideal solution [6]. TOPSIS steps are shown as below:

Step 1: Construct decision matrix, $\left(\left(x_{ij}\right)_{m\times n}\right)$

$$(x_{ij})_{m\times n} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ . & & & . \\ . & & & . \\ . & & & . \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \tag{11}$$

Step 2: Decision matrix normalization, $(r_{ij})_{m\times n}$

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \text{ where } i = 1,2,...,m \text{ and } j = 1,2,...,n \tag{12}$$

$$R = (r_{ij})_{m\times n} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ . & & & . \\ . & & & . \\ . & & & . \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix}$$

Step 3: Construct the weighted normalized decision matrix formation

$$\mathrm{T} = (t_{ij})_{m\times n} = (w_j r_{ij})_{m\times n}, i = 1,2,...,m, \sum_{j=1}^{n} w_j = 1$$

w_j is the equal weight that is given to the indicator where the w_j is equal to 0.1. This weight was determined by dividing 1 by the total of 10 financial indicators.

$$\mathrm{T} = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \dots & w_n r_{1n} \\ w_1 r_{21} & w_2 r_{22} & \dots & w_n r_{2n} \\ . & & & . \\ . & & & . \\ . & & & . \\ w_1 r_{m1} & w_2 r_{m2} & \dots & w_n r_{mn} \end{bmatrix} \tag{13}$$

Step 4: Determine the positive/ best ideal and the negative/ worst ideal solution
Beneficial indicators:

$$A_b = \{\langle \min(t_{ij} \mid i = 1,2,...,m) \mid j \in J_- \rangle, \tag{14}$$
$$\langle \max(t_{ij} \mid i = 1,2,...,m) \mid j \in J_+ \rangle\} \equiv \{t_{bj} \mid j = 1,2,...,n\},$$

$$A_b = \{\langle \max(t_{ij} \mid i = 1,2,...,m) \mid j \in J_- \rangle, \tag{15}$$
$$\langle \min(t_{ij} \mid i = 1,2,...,m) \mid j \in J_+ \rangle\} \equiv \{t_{bj} \mid j = 1,2,...,n\},$$

where,
$J_+ = \{j = 1,2,...,n \mid j$ associates with the criteria having a positive impact, and
$J_- = \{j = 1,2,...,n \mid j$ associates with the criteria having a negative impact.

Step 5: Calculate the separation measures

$$d_{ib} = \sqrt{\sum_{j=1}^{n} (t_{ij} - t_{bj})^2}, i = 1,2,...,m \tag{16}$$

$$d_{iw} = \sqrt{\sum_{j=1}^{n} (t_{ij} - t_{wj})^2}, i = 1,2,...,m \tag{17}$$

Step 6: Calculate the relative closeness to the ideal solution for each alternative solution, s_{iw}

$$s_{iw} = \frac{d_{iw}}{d_{ib} + d_{iw}}, 0 \le s_{iw} \le 1, i = 1,2,...,m \tag{18}$$

Step 7: Rank the alternatives

The result is then can be rank and analyzed.

Empirical Results

Table 2, Table 3 and Table 4, present the multi-criteria decision-making matrix, normalized decision matrix and weighted normalized decision matrix respectively. The best ideal solution and worst ideal solution for each decision criterion are presented in Table 5: Best ideal A_b and worst ideal A_w solution

Table 2: Multi-criteria decision-making matrix

	Return on equity	Return on assets	Net profit margin	CR	Debt to equity	Debt to assets	P/E ratio	EPS	Divide-nd yield	ART
C1	4.07	2.561	11.87	1.96	0.29	0.2	25.01	0.06	4.11	0.22
C2	8.87	2.98	6.87	1.29	0.68	0.23	7.42	0.08	4.7	0.43

C3	2.10	1.23	8.16	1.68	0.61	0.31	16.35	0.05	1.71	0.15
C4	2.70	1.5	5.25	1.2	0.67	0.39	19.93	0.04	1.83	1.05
C5	27.9	11.2	17.9	1.77	0.78	0.53	5.99	0.19	3.75	1.77

Table 3: Normalized decision matrix

	Return on equity	Return on assets	Net profit margin	CR	Debt to equity	Debt to assets	P/E ratio	EPS	Divide-nd yield	ART
C1	0.1366	0.2129	0.4835	0.5463	0.2048	0.2536	0.6730	0.2704	0.5336	0.1038
C2	0.2981	0.2478	0.2798	0.3595	0.4839	0.2916	0.1997	0.3606	0.6102	0.2029
C3	0.0706	0.1023	0.3324	0.4673	0.4341	0.3931	0.4400	0.2164	0.2220	0.7078
C4	0.0907	0.1247	0.2139	0.3338	0.4768	0.4945	0.5363	0.1848	0.2376	0.4955
C5	0.9377	0.9313	0.7292	0.4924	0.5551	0.6720	0.1612	0.8461	0.4869	0.8352

In Table 3, SkyWorld Development Berhad (C5) remains on the upward trajectory by scoring high on the normalized values. However, S P Setia Berhad (C1) has a strong score in account receivables turnover that is equal to 0.7078, which was an indicator of the efficiency of receivables, rest of the company such as Sime Darby Property Berhad (C1) and LBS Bina Group Berhad (C2) are moderate.

Table 4: Weighted normalized decision matrix

	Return on equity	Return on assets	Net profit margin	CR	Debt to equity	Debt to assets	P/E ratio	EPS	Divide-nd yield	ART
C1	0.0137	0.0213	0.0484	0.0546	0.0205	0.0254	0.0673	0.0270	0.0534	0.0104
C2	0.0298	0.0248	0.0280	0.0359	0.0484	0.0292	0.0200	0.0361	0.0610	0.0203
C3	0.0071	0.0102	0.0332	0.0467	0.0434	0.0393	0.0440	0.0216	0.0222	0.0071
C4	0.0091	0.0125	0.0214	0.0334	0.0477	0.0495	0.0536	0.0185	0.0238	0.0495
C5	0.0938	0.0931	0.0729	0.0492	0.0555	0.0672	0.0161	0.0846	0.0487	0.0835

The weighted normalized decision matrix in Table 4 indicates distribution of weights for each criterion according to their significance. SkyWorld Development Berhad (C5) is thus, the overall winner and corroborated by high weighted scores accorded for profitability such as the return on equity and return on assets. However, Sime Darby Property Berhad (C1) and LBS Bina Group Berhad (C2) rated moderately and do not lead in all the criteria.

Table 5: Best ideal A_b and worst ideal A_w solution

	Return on equity	Return on assets	Net profit margin	CR	Debt to equity	Debt to assets	P/E ratio	EPS	Divide-nd yield	ART
A_b	0.0938	0.0931	0.0729	0.0546	0.0205	0.0254	0.0673	0.0846	0.0610	0.0835

A_w	0.0071	0.0102	0.0214	0.0334	0.0555	0.0672	0.0161	0.0185	0.0222	0.0071

Table 6: Distance of the alternatives from the best and worst ideal solution

Company	(d_{ib})	(d_{iw})	S_{iw}
C1	0.1446	0.0894	0.3822
C2	0.1432	0.0654	0.3136
C3	0.1691	0.0451	0.2104
C4	0.1588	0.0600	0.2742
C5	0.0760	0.1680	0.6884

Table 7: Ranking of the financial performance of the property companies in Malaysia

Company	S_{iw}	Rank
SkyWorld Development Berhad (C5)	0.6884	1
Sime Darby Property Berhad (C1)	0.3822	2
LBS Bina Group Berhad (C2)	0.3136	3
Glomac Berhad (C4)	0.2742	4
S P Setia Berhad (C3)	0.2104	5

Table 7, presents the ranking of financial performance of Malaysian property companies. SkyWorld Development Berhad (C5) recorded the highest score (S_{iw} = 0.6884), followed by Sime Darby Property Berhad (C1), LBS Bina Group Berhad (C2), Glomac Berhad (C4), and S P Setia Berhad (C3) at the lowest rank. The findings suggest that firms ranked higher are closer to the positive ideal solution, reflecting stronger financial health and providing useful benchmarks for performance evaluation.

Conclusion

Evaluate company performance is vital for continuous improvement and informed decision-making. This study demonstrates that SkyWorld Development Berhad ranked highest among the selected Malaysian property companies, largely due to its superior profitability and balanced overall financial performance. In contrast, S P Setia Berhad ranked lowest, indicating weaker alignment with the ideal standard. The findings highlight the usefulness of the TOPSIS method as a replicable framework for benchmarking corporate financial efficiency and supporting investment appraisals. However, this study is limited to selected financial indicators within a specific timeframe and industry scope. Future research could incorporate broader financial and non-financial measures, as well as comparative analyses across different sectors, to enhance the robustness and applicability of the findings.

Acknowledgment

The authors thank Universiti Teknologi MARA for the facilities and support. Special thanks to Madam Farah Waheeda Binti Azhar for her guidance and support. Appreciation is also given to everyone who contributed to this research.

References

[1] Al Kharusi, S., & Başci, E. S. (2017). Financial institutions performance evaluation in a unique

developing market using TOPSIS approach. *Banks and Bank Systems*, *12*(1), 54–59. https://doi.org/10.21511/bbs.12(1).2017.06

[2] Azhar, F. W., Abd Rahim, Z. H., Fahami, N. A., Abdul Rahim, S. K. N., & Karim, H. A. (2022). Investing in Malaysian healthcare using technique for order preference by similarity to ideal solution. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(3), 1723–1730. https://doi.org/10.11591/ijeecs.v25.i3.pp1723-1730

[3] Banu, Dr. A. R. R., & Santhiyavalli, Dr. G. (2018). A Topsis approach to evaluate the financial performance of scheduled commercial banks in India. *International Journal Economics*, *9*(6), 71–92.

[4] Doka, K. M., Ahmad, F., Wan Shamsuddin, S. N., Wan Awang, W. S., & Ghazali, N. (2015). Integrated decision support system for human resource selection using TOPSIS based models. *Applied Mathematical Sciences*, *9*(129), 6403–6414. https://doi.org/10.12988/ams.2015.53288

[5] Heriyanto, S., Nuryanto, U. W., Pratiwi, I., & Prihatin, J. (2022). *Comparative Analysis Financial Performance of Property and Real Estate Sector Before and During Pandemic Covid-19 at Indonesia*. www.idx.co.id

[6] Hoe, L. W., Siew, L. W., & Fai, L. K. (2019). Performance analysis on telecommunication companies in Malaysia with TOPSIS model. *Indonesian Journal of Electrical Engineering and Computer Science*, *13*(2), 744–751. https://doi.org/10.11591/ijeecs.v13.i2.pp744-751

[7] Jan Mohd Khan, S., Nurazira Mohd Daud, S., Yushairi Mat Yusoff, M., Woei Chyuan, W., & Erman Che Johari, E. (2024). Property market and the Financial sector: Exploring Malaysia's scenario in times of crisis. In *Journal of the Malaysian Institute of Planners*, 22(1), 409-423, https://doi.org/10.21837/pm.v22i30.1449

[8] Kuangting Yang. (2022). Application of TOPSIS Technique for Financial Performance Evaluation of Manufacturing Firms in US Market. *Academic Journal of Business & Management*, *4*(16), 116-122, https://doi.org/10.25236/ajbm.2022.041620

[9] Roy, S., & Das, A. (2018). Application of Topsis method for financial performance evaluation: A study of selected scheduled banks in Bangladesh. *Journal of Commerce & Accounting Research*, 7(1), 24-29, http://www.publishingindia.com

[10] Sevkli, M., Zaim, S., Turkyilmaz, A., & Satir, M. (2010). An application of fuzzy TOPSIS method for supplier selection. *2010 IEEE World Congress on Computational Intelligence, WCCI 2010*. https://doi.org/10.1109/FUZZY.2010.5584006

[11] Zull Kepili, E. I., & Masron, T. A. (2016). Malaysia property sector: Performance analysis and portfolio diversification benefits within sub-sector. *International Journal of Housing Markets and Analysis*, *9*(4), 468–482. https://doi.org/10.1108/IJHMA-08-2015-0043

MUD CRAB WEIGHT PREDICTION USING MULTIPLE LINEAR REGRESSION MODEL

Yusrina Andu[1,*] and Maisarah Saqinah Mohd Ridzwan[2]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kampus Kuala Pilah, 72000 Kuala Pilah, Negeri Sembilan*
[2]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan, Malaysia.*
*yusrinaandu@uitm.edu.my

Abstract

This study explores the use of multiple linear regression (MLR) to predict the weight of mud crabs (Scylla spp.) based on gender, carapace length (CL), sternum width (SW), and propodus width (PW). *Data from 60 mud crabs comprising two species, Scylla paramamosain and Scylla olivacea, were analyzed. Key predictors, including CL, SW, and PW, demonstrated significant correlations with body weight, explaining 95.9% of the variance in the final regression model. Model validation showed high reliability, with adherence to linearity between the variables, normality, homoscedasticity, and error terms assumptions. The findings offer a practical tool for fisheries management, allowing efficient and non-invasive weight estimation, which aids in sustainable harvesting and population assessments. Despite the model's strong predictive capabilities, the study acknowledges limitations, such as reliance on secondary data and geographic constraints, and suggests the inclusion of additional variables and expanded datasets to enhance generalizability. This research provides a foundation for improved aquaculture practices and the development of adaptive management strategies in dynamic environmental conditions.*

Keywords: Mud Crabs, Morphometric Measurements, Multiple Linear Regression

Introduction

Mud crabs (*Scylla sp.*) are a valuable aquatic resource with high economic value, making them an important species for fisheries management. Mud crabs are large crabs found in estuarine and mangrove habitats throughout the Indo-Pacific region, especially Malaysia [1-3] These crabs are highly valued for their meat, making them an important species in both commercial fisheries and aquaculture. There are several species of mud crabs, with the most recognized being Scylla serrata. The management of mud crab fishery ultimately depends on the understanding of the factors influencing their population dynamics, growth patterns, and size distribution [2][4]. Furthermore, the length-weight relationship of mud crabs has been considered the most important foundation in fish biology, the knowledge of which can provide insight into the general health and productivity of crab populations. Additionally, considering the influence of gender on crab growth and size distribution is essential for implementing targeted management strategies.

Mud crab (*Scylla sp.*) also is one of the waters resources, which can have high economic value. Therefore, scientific information related to length-weight relationships is necessary for sustainable fisheries management [5]. The length-weight relationship is one of the paramount things which must be considered before the fish binomial studies. In this viewpoint, the relationship is commonly used to index aquatic life including fish and crustaceans [6]. It can discover the variation caused by this odd length and can be used as an indicator of the fatness, breed of the species, and feed of the species and the way the species has adapted to the environment and if it is optimal or not. Data that tempers its population biology in the Malaysian country which relates to its population biology in Malaysia is limited and lacks depth. The research on the length-weight relationship in crustaceans including mud crabs has come into the spotlight by virtue of the physiological processes and the structure and function of tissues involved in fish. It has been used as an indirect tool to study population dynamics, and growth rates. This is because knowledge on how crab weight range with length is essential to stock assessment of biomass, size and stock status estimation. Besides that, through the gender integration gender in the analysis allows us to study whether the male and the female crabs presents different growth patterns and size distributions which are very important in the fisheries management practices and sustainability [4].

Morphometric measurements can be defined as the process of measurement of various features of an organism and quantifying a creature's shape. These measurements are used to describe the size and shape such as length, width, height and even circumference of any body parts of the creature. For instance, in the analysis of mud crabs, the morphometric measurements may be limited to, the carapace length and width, sternum width, abdomen width, and length and width of propodus that is part of a claw. These measurements are critical in biosciences especially when defining the differences between one species over the other, monitoring growth rates, and defining the morphology of the physical properties of an organism that may define its functionality and behavior in the ecosystem. Morphometric data may also be incorporated into other statistical models for the estimate of other light biological characteristics, such as weight, from these dimensions.

Methodology

The methodology for this study was designed to predict the weight of mud crabs using multiple linear regression (MLR). The research relied on secondary data comprising 60 observations of two mud crab species, *Scylla paramamosain* and *Scylla olivacea*, sourced from Kelantan, Malaysia. The data included morphometric measurements such as carapace length (CL), sternum width (SW) and propodus width (PW), along with gender as categorical predictors. The dependent variable was body weight, expressed in grams. The primary objective of the study was to develop a model for predicting mud crab weight using MLR and to identify significant factors influencing weight.

Theoretical Framework

Figure 1 illustrates that the body weight of mud crabs, the dependent variable in this study, is influenced by the independent variables, which are the morphometric measurements: CL, SW

and PW.

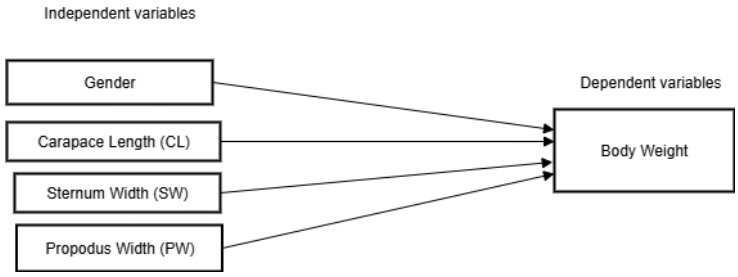


Figure 1: Theoretical Framework

Multiple linear regression

Multiple linear regression (MLR) is a statistical technique that represents the relationship between one dependent variable and two or more independent variables. The model assumed a linear relationship between the dependent variable and each of the independent variables, as shown in Eq.1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \quad (1)$$

where,

Y : Body weight

β_0: y-intercept

β_p: Coefficient for each independent variables

X_1: Gender

X_2: Carapace Length (CL)

X_3: Sternum Width (SW)

X_4: Propodus Width (PW)

ε: Error Term

Result and Analysis

By using MLR, we can determine the measurements that affect the body weight of mud crab. The result is present in Table 1.

Table 1: ANOVA Table

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	12.501	4	3.125	325.914	<0.001
Residual	0.489	51	0.010		
Total	12.990	55			

Based on Table 1, the significant value is < 0.001 which is less than α value equal to 0.05, therefore, it can be concluded that the model is fit and significant.

Table 2: Model Summary

R Square	0.962
Adjusted R^2	0.959

It can be observed from Table 2, R^2 value is 0.962, this indicates that about 96.2% of total variation in body weight of mud crab is explained by gender, CL, SW and PW, while the balance 3.8% is explained by other factors. A high R^2 value, close to 1, suggests a very good fit, meaning the model explains most of the variability in the dependent variable. This implies that these predictors are highly influential in determining the weight of the mud crab. Additionally, it is also important to consider the Adjusted R^2, as it accounts for the number of predictors and the sample size. In this case, the Adjusted R^2 is 0.959, which is very close to the R^2, suggesting that the inclusion of predictors has not led to overfitting.

Table 3: Coefficients of Significant Regression Table

Model	**B**	**t**	**Significance**
(Constant)	2.190		
Gender	– 0.100	-2.740	0.008
Carapace Length (CL)	0.020	3.600	<0.001
Sternum Width (SW)	0.023	6.807	< 0.001
Propodus Width (PW)	0.023	3.325	0.002

All these four variables affect the body weight of mud crab since they all have significant value less than α value equal to 0.05 (Table 3). The final MLR model is as follows:

$$\hat{Y} = 2.190 - 0.100 \cdot \text{Gender} + 0.013 \cdot \text{CL} + 0.037 \cdot \text{SW} + 0.024 \cdot \text{PW}$$

When carapace length, sternum width and propodus width are not available or at zero, the average of body weight of mud crab will lie at 2.190. Meanwhile on average, when gender increase by one unit, body weight of mud crab will decrease by 0.100, while propodus width and carapace width is held constant. In this model, the assumptions of MLR which are linearity, normality, homoscedasticity, and independence of error terms were all satisfied.

Conclusion

This study successfully developed a predictive model for mud crab body weight as a function of morphometric measurements; in which carapace length and sternum width were the most significant predictors. Additionally, gender was identified as a contributory factor, but less so than in the case of other variables. The best MLR model showed excellent predictability, as indicated by an R^2 of 0.962, and an adjusted R^2 of 95.9% of the variation in body weight. The

results also highlighted the importance of considering several variables when predicting stock depletion, which could help scientists and fisheries managers better manage fishing stocks. This model will help simplify the estimation of mud crab weight through simple measurements for resource assessment to ensure sustainable exploitation. Furthermore, there were strong positive correlations between morphometric characteristics and weight, confirming the hypothesis that these physical traits are good indicators of the crab weight. This adds validity to the potential utility of the model in real-world settings as well as providing a scientifically sound framework for population monitoring and ecological assessment.

References

[1]Azra, M. N., & Ikhwanuddin, M. (2016). A review of maturation diets for mud crab genus Scylla broodstock: Present research, problems and future perspective. *Saudi Journal of Biological Sciences*, *23*(2), 257-267.

[2]Pane, A. R. P., Kembaren, D. D., & Suman, A. (2021, October). The fishing season and the exploitation status of mud crab (Scylla serrata) in Asahan Sea, North Sumatera, Indonesia. In *IOP Conference Series: Earth and Environmental Science* (Vol. 890, No. 1, p. 012045). IOP Publishing.

[3]Waiho, K., Fazhan, H., & Ikhwanuddin, M. (2016). Size distribution, length–weight relationship and size at the onset of sexual maturity of the orange mud crab, Scylla olivacea, in Malaysian waters. *Marine Biology Research*, *12*(7), 726-738.

[4]Ikhwanuddin, M., Azmie, G., Juariah, H. M., Zakaria, M. Z., & Ambak, M. A. (2011). Biological information and population features of mud crab, genus Scylla from mangrove areas of Sarawak, Malaysia. *Fisheries Research*, *108*(2-3), 299-306.

[5]Nurfadillah, N., Silvia, N., Octavina, C., Dewiyanti, I., & Rizwan, T. (2021, February). Length-weight relationship (LWR) of mud crabs (Scylla sp.) in mangrove waters Peukan Bada Aceh Besar as the basic for waters resources development. In *IOP Conference Series: Earth and Environmental Science* (Vol. 674, No. 1, p. 012068). IOP Publishing.

[6]Khan, M. A., & Mustaqeem, J. (2013). Carapace width weight relationship of mud crab Scylla serrata (Forskal, 1775) from Karachi Coast. *Canadian Journal of Pure and Applied Sciences*, *7*(2), 2381-2386.

Evaluating Students' Level of Financial Management: A Case Study

Nora Mohd Basir[1]*, Nurafrina Batrisyia Johari[1]
[1] *College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, 70300 Seremban Negeri Sembilan, Malaysia*
*noram661@uitm.edu.my

Abstract

Financial management is crucial for students as it helps them manage their money effectively, ensuring they can cover essential expenses like tuition, books, and living costs. In order to ensure future financial stability, it also imparts essential skills such as debt management, investing, and saving. By engaging in sound financial management practices, students can reduce stress, avoid unnecessary debt, and lay the groundwork for long-term financial well-being. Thus, this study attempts to identify students' level of financial management. 398 students from one of the public universities in Negeri Sembilan were selected as a sample to answer the online survey. The analysis results show that 78.6% of the students are classified as manageable in managing finances and 21.4% as unmanageable. Hence, it is hoped that students' financial management awareness will increase and they can achieve better stability and security throughout their lives.

Keywords*:* Financial management, Financial stability, Public universities, Students

Introduction

Financial management is an essential life skill everyone must master, especially in today's rapidly evolving financial landscape. Technological advancements have transformed financial services and products, making financial decision-making more complex. The vast array of financial products and global economic uncertainties require consumers to navigate increasingly complicated financial choices. The rapid growth of technology has expanded the role of financial services and enhanced accessibility to evolving tools such as online and mobile banking. Additionally, consumers can now purchase goods and services more easily and quickly than ever. However, individuals must develop strong financial literacy and management skills to make informed financial decisions in this digital era.

Achieving financial stability and success requires effective financial management, which involves planning, budgeting, and exercising financial control. This skill is particularly crucial for students in higher education as they transition into adulthood and experience financial independence for the first time. This transition is vital to managing expenses, handling debt, and making responsible financial choices. Mastering these skills impacts students' academic experiences and lays the foundation for their long-term financial well-being. Money plays a crucial role in everyday life, as many needs and responsibilities cannot be met without it. While some may argue that money is not the key to happiness, it remains a fundamental necessity for nearly all aspects of life. Students often experience financial pressures stemming from increasing tuition fees, living expenses, and the necessity of educational loans. According to [1] the demands of life and current financial needs in higher education institutions can be quite

challenging. As financial demands increase, applications for financial aid have risen, reflecting the growing economic challenges students face.

University and college students, in particular, must develop sound financial management habits, as their financial behavior has long-term implications for their post-graduation lives. To help students balance their needs with the available financial resources, students must be financially literate and skilled [4]. Poor financial management can lead to high debt levels, academic struggles, and psychological stress, whereas strong financial skills empower students to make informed choices, achieve financial goals, and maintain a balanced lifestyle throughout their academic journey.

Given the growing financial pressures on students, financial management has become a critical area of study. Students must have financial knowledge and skills to align their needs with available financial resources [4]. Many students tend to overspend on non-essential items, such as branded products and luxury goods, without considering the long-term consequences. Additionally, a significant number of students take on part-time jobs to supplement their income and afford off-campus living expenses [2]. Effective money management enables individuals to make sound financial decisions that promote resilience, stability, and growth. Budgeting, responsible spending, and debt avoidance are key aspects of financial well-being. For instance, a student who carefully prepares a monthly budget, tracks income from a part-time job, and allocates funds for tuition, rent, and groceries can maintain financial stability and avoid unnecessary debt. Without such financial discipline, students risk accumulating high-interest credit card debt, which can become increasingly difficult to repay over time, ultimately jeopardising their financial security.

Thus, this study aims to assess students' financial management levels and provide insights that can help students, educators, universities, policymakers, and financial institutions develop targeted interventions. The findings will enhance financial literacy and promote better financial management practices among students, ensuring their long-term financial success.

Methodology

Students from a public university in Negeri Sembilan were selected as the target population for this study. A simple random sampling method was employed, as a sampling frame of students was accessible from the university administration. A total of 398 questionnaires were distributed to the chosen students via Telegram. The questionnaire was adapted from [3] and organised into two sections. Section A aimed to collect demographic information from respondents, including gender, primary source of finance, monthly income, and savings. Section B evaluated students' financial management skills through seven questions using a 5-point Likert scale. The data was analysed using descriptive analysis. A binary result derived from respondents' overall scores on the financial management questionnaire where a total score between 9 and 21 is divided into" unmanageable", denoting less successful money management techniques. The manageable, which reflects better financial management techniques, is made up of respondents with scores between 22 and 35. Therefore, this study used a table to provide clear descriptions of datasets, identify patterns, gain insights into the information, and effectively communicate key variables.

Results and Discussion

After a thorough data collection, the responses of 398 participants were analysed. This analysis was crucial in constructing a clear and detailed profile of the respondents, highlighting their individual characteristics to uncover significant insights and trends within the data effectively.

Table 1: Descriptive Information

	Frequency	%
Gender		
Female	306	*76.9*
Male	92	*23.1*
Savings		
No	108	*27.1*
Yes	290	*72.9*
Main Source of Finance		
Family	170	*42.7*
Scholarship	42	*10.6*
Self-Finances	13	*3.3*
Study Loan	173	*43.5*
Total	398	*100.0*

Table 1 shows that 76.9% of the respondents are female, with the remaining 23.1% being male. Moreover, 72.9% of respondents allocate a portion of their income to monthly savings, while 27.1% do not. Regarding financial sources, most students depend on study loans and family support, accounting for 43.5% and 42.7%, respectively. Meanwhile, scholarships provide funding for 10.6% of students, and 3.3% rely on self-financing as their primary source.

In examining student financial management, the five-point Likert scale (Strongly disagree, Disagree, Neutral, Agree and Strongly agree) was divided into two groups for analysis. Disagree and Strongly Disagree responses were grouped as less proactive financial management (Unmanageable), while Neutral, Agree, and Strongly Agree were merged and classified as proactive financial management (Manageable). This method makes it easier to compare students who actively manage their finances and those who are less engaged. Table 2 indicates that 67.9% of respondents demonstrate a proactive approach and a strong commitment to controlling their spending, while 32.2% are less engaged. Most people prioritize saving money, with 69.1% actively putting money aside, while 30.9% are less likely to do so. Moreover, 40.7% of the respondents do not prioritize this habit, while 59.3% of the respondents proactively plan their daily expenses. Another sign of financial discipline is creating a careful monthly budget, which 57.1% of the proactive group and 42.9% of the less proactive group adopt. 59.5% of the respondents said that they track their purchases to keep costs down, while 40.5% said that they do so less frequently. The most proactive shopping behaviour is comparing prices of goods when shopping, with 76.6% of consumers doing so, compared to only 23.3% who do not. Lastly,

64.9% of people have a habit of planning for unforeseen costs, whereas 35.1% are less ready for unforeseen financial circumstances.

Table 2: Percentage of Financial Management Statement

Financial Management	Level of Agreement					Total Row
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	
	%					
I spend money according to the budget that I have made.	*0.8*	*7.8*	*23.6*	*46.5*	*21.4*	*100.0*
I set aside some of my money for savings	*2.8*	*6.5*	*21.6*	*48.2*	*20.9*	*100.0*
I plan my daily expenses.	*2.8*	*11.1*	*26.9*	*41.7*	*17.6*	*100.0*
I will follow a careful monthly budget.	*3.0*	*11.3*	*28.6*	*40.5*	*16.6*	*100.0*
I always take note of every purchase I make to limit the expenses	*2.8*	*13.8*	*23.9*	*42.7*	*16.8*	*100.0*
I always compare the prices goods when shopping.	*0*	*7.0*	*16.3*	*42.7*	*33.9*	*100.0*
I prepare for unexpected expenses	*3.3*	*8.5*	*23.4*	*40.5*	*24.4*	*100.0*

Table 3 provides a summary of students' financial management levels at public universities in Negeri Sembilan, addressing the objective of identifying how well students manage their finances. Among the 398 students, 313 (or 78.6%) indicated that they could handle their finances effectively. In contrast, the remaining 21.4% reported difficulties in managing their finances. Overall, the study findings reveal that most of the students demonstrated effective financial management skills.

Table 3: Level of Students' Financial Management

Level	Frequency	%
Unmanageable	85	*21.4*
Manageable	313	*78.6*
Total	398	*100.0*

Conclusion and Recommendations

The assessment aimed at determining students' financial management skills revealed that the majority of students handle their finances wisely. Specifically, 78.8% are classified as "manageable" in their financial behavior, which suggests a positive outcome. This is an excellent result. However, it is important to acknowledge that a subset of students, 21.4%, are deemed "unmanageable" and appear to struggle with effectively managing their money. This research emphasises the importance of targeted educational programs and awareness campaigns to address the knowledge gap in financial management and ensure a better understanding of its impact on future income stability. Therefore, the responsible persons, particularly the educator and the university, should collaborate with financial institutions to lead an awareness campaign to enhance students' understanding of money management. Ultimately, effective financial management empowers students to achieve their goals and secure a stable future.

To thoroughly explore students' financial management skills, future research should be conducted on a larger scale, extending beyond just university students. Studying a diverse group of students offers valuable insights into financial management dynamics across different demographics. A comprehensive understanding requires a broad and varied sample size that includes individuals from various groups and socioeconomic backgrounds. This inclusive approach ensures a more representative and practical analysis of the wider population's financial management capabilities. By expanding the focus beyond just students, researchers can obtain more accurate evidence of societal skills in managing finances. This, in turn, can aid in developing targeted interventions and educational campaigns that address the different perspectives and experiences related to financial management across various segments of the population.

Acknowledgments

We sincerely thank each member for their unwavering support and commitment to ensuring this article was published successfully.

References

[1] Barr, M. J., & McClellan, G. S. (2018). Budgets and financial management in higher education. John Wiley & Sons.

[2] Douglas, D., & Attewell, P. (2019). The relationship between work during college and post college earnings. Frontiers in Sociology, 4, 78.

[3] Oseifuah, E. K. (2014). Analysis of the level of financial literacy among South African undergraduate students. Journal of Economics and Behavioral Studies, 6(3), 242– 250.

[4] Zulfaris, M. D., Mustafa, H., Mahussin, N., Alam, M. K., & Daud, Z. M. (2020). Students and money management behavior of a Malaysian public university. The Journal of Asian Finance, Economics and Business, 7(3), 245–251.

NUMERICAL SOLUTION FOR MHD FLOW OVER A SHRINKING WEDGE USING BVP4C

Nurul Umairah Ahmad[1,], Siti Hidayah Muhad Saleh[2]
[1]*College of Computing, Informatics and Mathematics, Universiti Teknologi Mara, Kampus Seremban, Negeri Sembilam*
[*]umyrahnrul@gmail .com

Abstract

This study investigates MHD flow over a shrinking wedge using BVP4C. The findings were significant for several industrial applications in MHD pumps, liquid metal processing, cooling and heating systems and aerodynamic designs. This study aims to find an alternative method for solving the MHD flow problem over contracting wedges. The objective includes to solve MHD flow over a shrinking wedge using BVP4C. The governing partial differential equations (PDE) for heat transfer and flow are converted to a system of ordinary differential equations (ODE) using similarity transformation. BVP4C in MATLAB software is then used to solve this equation numerically. The study then compares the result between shooting method and BVP4C, thus analyse the effect of shrinking parameter, λ *and Hartree pressure gradient,* β *on velocity profile, temperature profile, skin friction coefficient and local Nusselt number. The comparison shows good agreement between the compared value of skin friction coefficient and local Nusselt number obtained in BVP4C. The result presents that an increase in* λ *will raise the velocity profile and decreased the temperature profile. There is also a second solution for cases of* $\lambda = -1.30, -1.31, -1.3994$ *when* $\beta = 0.1$.

Keywords*:* Magnetohydrodynamic, shrinking wedge, BVP4C

1. Introduction

This study explores boundary layer flow in various daily life situations, particularly in industries. Boundary layers are small fluid layers that develop on a bounding surface as fluid moves. The fluid interacts with the wall, creating a non-slip boundary region condition. The study also explores Magnetohydrodynamics, which involves the flow of electrically conducting fluids through a magnetic field. This fluid flow have practical applications in various technological processes, including plasma research, petroleum industry, power generator design, nuclear reactor cooling, and heat exchanger construction [1]. The flow of electrically conducting fluids over an external voltage-driven electric current or magnetic field is known as magnetohydrodynamics [2], and it is covered in this subject. This study focuses on MHD fluid flow towards shrinking wedges, a topis less explored in prior research.. Heat transfer is included in the flow process of the investigation, which may affect the methodology and findings. Convection, conduction, and thermal radiation are additional parameters included in the study.

According to [7], a wedge is a simple machine with a triangular shape and a sloping surface that is thicker at one end and narrows to a sharp edge at the other. A diminishing wedge is a

situation in which the inclined planes converge to a single point. Normally, a wedge is constructed by consisting of one or two inclined planes. The flow towards a wedge surface in fluid dynamics involves several factors, including partial slip and heat transfer, which can have a big impact on the methodology and findings of the study. Since shrinking wedge flow is the subject of this investigation, thermal energy transfer occurs naturally as the fluid passes over the wedge surface. The several ways that energy and entropy are transferred from one area to another are referred to as heat transfer [11].

The study about MHD boundary layer flow of a nanofluid across a dynamic stretched surface by using Rungge-Kutta methods was conducted by [3]. The results of the study showed that the viscoelastic and magnetic parameters affect the non-dimensional velocity, temperature and concentration profile, where there is an increase in the velocity while the temperature decreased. Meanwhile, [2] used Runge-Kutta Fehlberg (RFK45) to examine the heat transfer process of MHD Powell-Eyring fluid and the mass via an absorptive flat surface while considering the radiation and mass diffusion effects. A study used SQLM to examine a wedge-shaped MHD boundary layer and a continuous flow of nanofluid submerged in porous media with viscous dissipation was conducted by [1]. A study by [4] analyzed the effects of a magnetic field on the mass and heat transfer of a liquid form over an unstable stretched cylinder using BVP4C. The curvature parameter increased the temperature and velocity profiles, while the instability parameterreduced the fluid thickness. A study implementing the spectral quasi linearization method (SQLM) was carried out to analyze stable laminar MHD flow in nanofluid over a wedge in the presence of changing magnetic fields [13]. The study found that nanoparticle volume fractions affected heat transfer in wedge shaped fluids with different angles and caused an increase in thermal conductivity and temperature.

Whilst [5] studied the flow of a thick, incompressible fluid in a magnetic field through a continuously stretched sheet in two dimensions. Using BVP4C, the study solved the boundary value problem and discovered that an increase in the thermal slip parameter (γ) it will decreased the skin friction and heat transfer rate, while the increase in magnetic parameter (M) it will increased the heat transmission. When the velocity parameter (δ) increased, the rate of heat transfer at the surface decreased. The heat transfer from MHD mixed convection flow to a penetrable stretched wedge implementing thermal radiation and ohmic heating using BVP4C was studiedby [12]. The numerical study showed that the skin friction, velocity, temperature and Sherwood number will all increased as the Hartree pressure gradient parameters increased. A study implementing the spectral quasi linearization method (SQLM) was carried out to analyze stable laminar MHD flow in nanofluid over a wedge in the presence of changing magnetic fields [13]. The study found that nanoparticle volume fractions affected heat transfer in wedge shaped fluids with different angles and caused an increase in thermal conductivity and temperature.

Thus, the focus of this study is to find a different approach to solving the MHD flow problem across contracting wedges while investigating it with MATLAB's BVP4C.

2. Methodology

Figure 1 represents the coordinates system of shrinking wedge and its physical model in 2-Dimensional Cartesian coordinates, x and y along the surface with a parameter u and v represent the velocity variables respectively. Assuming the shrinking wedge velocity is $u_w(x) = U_w x^m$, where $U_w < 0$ represents the wedge is shrinking, while m and U^∞ are illustrated as a positive constant. A variable magnetic field, $B(x)$ with assuming a small induced magnetic field and will not be considered is applied along the positive y-axis. The Hartree pressure gradient parameter, β corresponds to $\beta = \frac{\Omega}{\pi}$ for a total angle, Ω of the wedge. $0 \leq m \leq 1$ is the interval of the boundary layer flow for a case that across the stationary flat plate in Blasius problem where $m = 0$ represent the boundary layer of the flat plate and $m = 1$ is for the flow towards a stagnation point in the boundless wall [6].

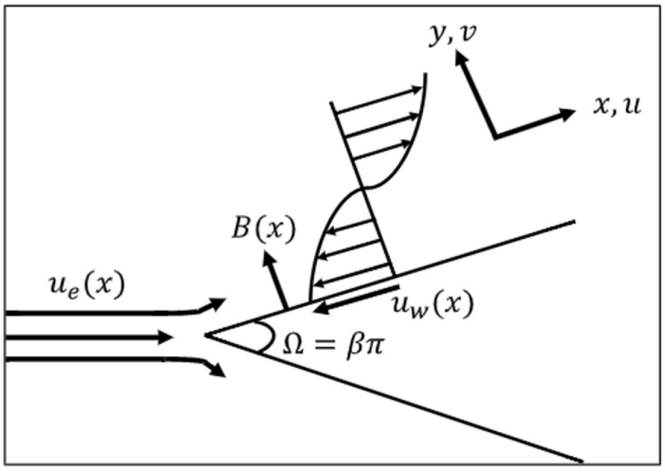


Figure 1: Coordinate system and physical model of shrinking wedge

Continuity equation, describing the conservation of mass in the boundary layer, can be expressed as

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{1}$$

Momentum equation, describing the balance of forces in the boundary layer, can be expressed as

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = u_e\frac{\partial u_e}{\partial x} + \nu\frac{\partial^2 u}{\partial y^2} - \frac{\sigma B^2(x)}{\rho}(u - u_e). \tag{2}$$

Energy equation governed by,

$$u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}=\alpha\frac{\partial^2 T}{\partial y^2}, \tag{3}$$

where u is a velocity for x-axis, v is a velocity for y- axis, u_e is velocity of external flow, $B(x)$ is the magnetic field, T is fluid temperature, α is thermal diffusivity, ρ is the fluid density and β is the Hartree pressure gradient parameter.

While the boundary conditions are

$$v=v_w,\quad u=u_w(x),\quad T=T_w\ \ at\ y=0 \tag{4}$$

$$u\to u_e(x),\quad T\to T_\infty\ as\ y\to\infty. \tag{5}$$

The similarity equation variables given by

$$\psi=\sqrt{\frac{2vxu_e}{1+m}}f(\eta),\ \eta=\sqrt{\frac{(1+m)u_e(x)}{2vx}}y,\ \theta(\eta)=\frac{T-T_\infty}{T_w-T_\infty}, \tag{6}$$

which u and v are defined as,

$$u=\frac{\partial\psi}{\partial y},\ v=-\frac{\partial\psi}{\partial x} \tag{7}$$

That satisfies the continuity equation in (1) and obtained as

$$u=u_e(x)f'(\eta), \tag{8}$$

$$v=-\sqrt{\frac{(1+m)vu_e(x)}{2x}}\left[f(\eta)+\frac{m-1}{m+1}\eta f'(\eta)\right], \tag{9}$$

where the prime represents the differentiation with respect to η, when the boundary $\eta=0$, then the transpiration rate is given by,

$$v_w=-\sqrt{\frac{(1+m)vu_e(x)}{2x}}s, \tag{10}$$

with $s = f(0)$, where the parameter $s > 0$ is for suction cases and $s < 0$ is for injection cases. The variables can be defined as ν is the kinematic viscosity, T_w is the uniform surface temperature, σ is electrical conductivity, ψ is stream function, $y_w(x)$ is the mass flux velocity, T_∞ is a free stream temperature and η is the boundary layer thickness.

Additional constraints, referred to as boundary conditions, must be applied to guarantee that the solutions of PDEs or ODEs satisfy requirements at the domain boundaries and to identify a single solution that matches the physical problem. At certain locations along the edge of the spatial region where the differential equation is held, the solution must meet certain requirements known as boundary conditions. The equations in this study indicate five boundary conditions as stated (4) and (5), where λ is defined as shrinking parameter, Pr is defined as the Prandtl number and s is define as the suction or injection parameter as below:

$$\lambda = \frac{U_w}{U_\infty}, \beta = \frac{2m}{m+1}, M = \sqrt{\frac{2\sigma}{(m+1)\rho U_\infty}} B_0, Pr = \frac{\nu}{\alpha}. \tag{11}$$

where U_w is a velocity and $\lambda < 0$ is both condition for shrinking case. Since there is no magnetic field, $M = 0$.and when $\beta = 0$ and $\beta = 1$, (12) will reduces to the Blasius and Hiemenz equations respectively. The PDE will be transformed by using a similarity transformation where the differentiated variable will be used to substitute into the PDE in (1), (2) and (3) to obtain the ODE as follows

$$f''' + ff'' + \beta\left(1 - f'^2\right) + M^2(1 - f') = 0, \tag{12}$$

and

$$\frac{1}{Pr}\theta'' + f\theta' = 0, \tag{13}$$

which subject to the boundary condition below

$$f(0) = s, f'(0) = \lambda, \theta(0) = 1. \tag{14}$$

$$f'(\eta) \to 1, \theta(\eta) \to \infty \; as \; \eta \to \infty. \tag{15}$$

This study only focuses on two physical quantities which is local Nusselt number, Nu_x and skin friction coefficient, C_f which both of it can be written as equation below:

$$C_f = \frac{\tau_w}{\rho {u_e}^2(x)}, \qquad Nu_x = \frac{x q_w}{k(T_w - T_\infty)}, \tag{16}$$

Then wall shear stress along a shrinking surface, τ_w and surface heat flux, q_w can be expressed as

$$\tau_w = \mu \left(\frac{\partial u}{\partial y}\right)_{y=0}, \quad q_w = -k \left(\frac{\partial T}{\partial y}\right)_{y=0}. \tag{17}$$

By substituting (17) in (16), the equation below was obtained

$$Re_x^{\frac{1}{2}} C_f = \left(\frac{1+m}{2}\right)^{\frac{1}{2}} f''(0), Re_x^{-\frac{1}{2}} Nu_x = -\left(\frac{1+m}{2}\right)^{\frac{1}{2}} \theta'(0). \tag{18}$$

3. Result and discussion

The results are compared with two previous findings by [6], who applied the shooting method with Rugge-Kutta Fehlberg (RFK45) in Maple software and [7], who applied a BVP4C method in MATLAB. The purpose of this comparison is to verify that the outcomes produced by BVP4C are validated. Table 1 represents the comparison results for skin friction coefficient, $f''(0)$, when $M = 0,\ \lambda = 0$ and $s = 0$ for different value of β. While Table 2 shows the comparison results for local Nusselt number when $M = 0,\ \lambda = 0$ and $s = 0$ for different value of β.

Table 1: The comparison results of skin friction coefficient when $\lambda = 0$ for different value of β

M	s	β	Skin friction coefficient $(f''(0))$		
			[7]	[6]	Present study
0	0	0	0.469600	0.469600	0.469600
		0.5	0.927680	0.927680	0.927680
		0.7	1.059808	1.059808	1.059808

Table 2: The comparison results of local Nusselt number when $\lambda = 0$ for different value of β

M	s	β	Local Nusselt Number, Nu_x	
			[6]	Present study
0	0	0	-0.469600	-0.469600
		0.5	-0.538978	-0.538979
		0.7	-0.533660	-0.553661
		1	-0.570465	-0.570465

From Table 1 and Table 2, it shows a good agreement of compared value for both skin friction coefficient and local Nusselt number result between the present study with the previous study where the value obtained by using BVP4C is similar as a value obtained from previous study using Maple.

3.1 Effect of shrinking parameter, λ on velocity profile, temperature profile, skin friction coefficient and local Nusselt number.

Figures 3.1 and Figure 3.2 present the velocity profiles, $f'(\eta)$ for various values of the shrinkage parameter, λ, where $M = 0.2,\ s = 1$ with Hartree pressure gradient parameter, $\beta = 0.1$ and $\beta = 0.33$. As shown in Figure 3.1, the increase in λ contributes to an increase of $f'(\eta)$. A decrease in λ reduces $f'(\eta)$ in Figure 3.2 while the boundary layer thickness is increasing when value of λ decrease and the value of Hartree pressure gradient, β increases.

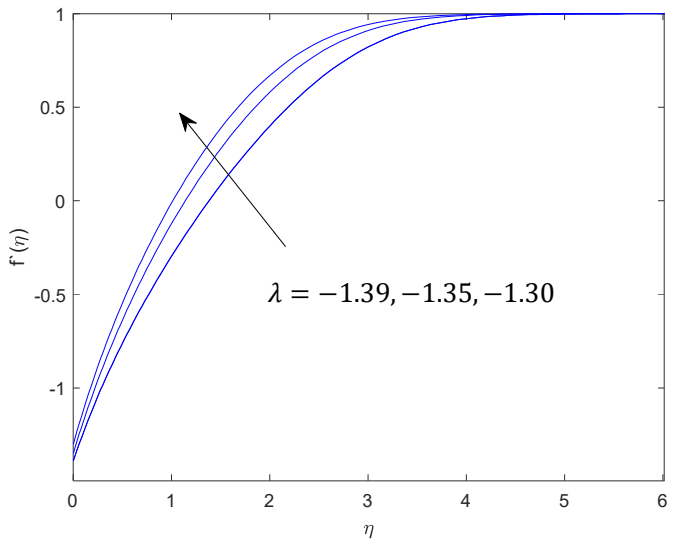


Figure 3.1: The velocity profile for different value of λ, when $\beta = 0.1$

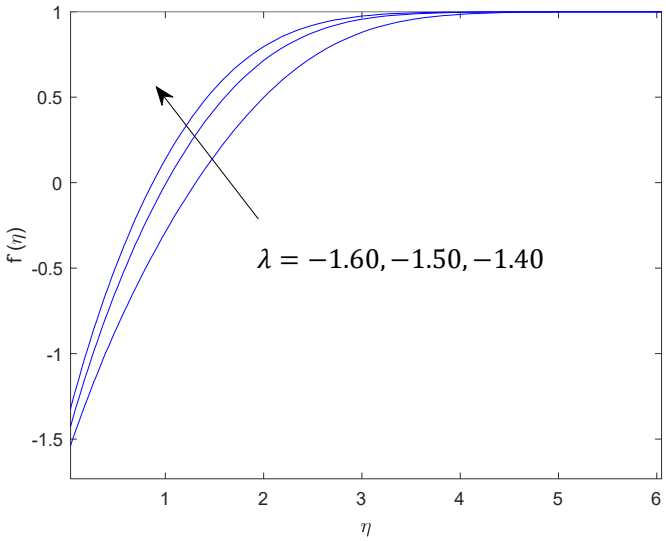


Figure 3.2: The velocity profile for different value of λ, when $\beta = 0.33$

Thus, the temperature profile for various values of λ when $M = 0.2,\ s = 1, Pr = 1$ with $\beta = 0.1$ and $\beta = 0.33$ is shown in the figure below, as in Figures 3.3 and Figure 3.4 respectively. The temperature profile decreases as the value of λ rises, as seen in Figure 3.3. This is due to the capacity of viscous dissipation to raise temperature. As the value of λ increases, the temperature profile decreases, and the boundary layer thickness is high as shown in Figure 3.4 when the value of Hartree pressure gradient parameter, β is increasing.

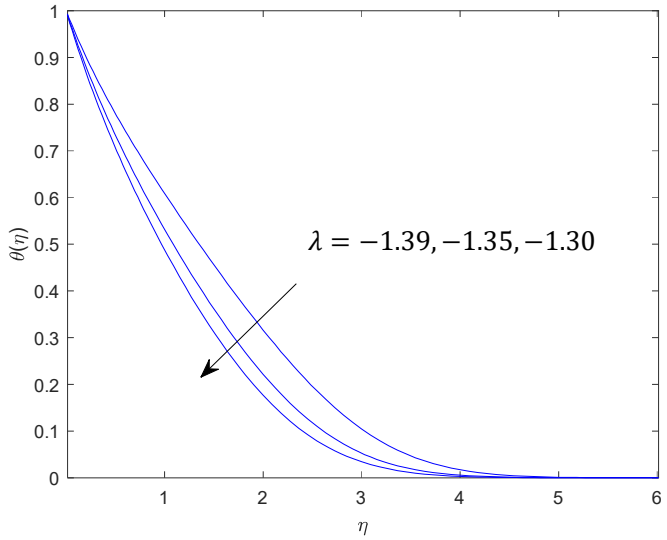


Figure 3.3: The temperature profile for different value of λ, when $\beta = 0.1$

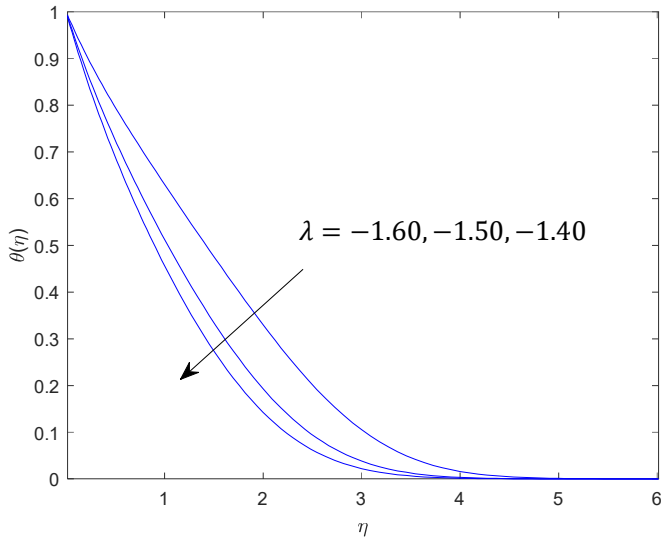


Figure 3.4: The temperature profile for different value of λ, when $\beta = 0.33$

Figure 3.5 and Figure 3.6 reveals the existence of the dual solution, subject to the boundary condition for a certain range of λ which depends on the skin friction coefficient and local Nusselt number. For the cases in this study, which is a shrinking wedge, $\lambda < 0$, the study presents the dual solution for the cases. Figure 3.5 illustrates the relation between λ and skin friction coefficient, $f''(0)$. It shows that the increase in value λ will decrease the value of $f''(0)$. Figure 3.6 shows the relation between λ and local Nusselt number, $\theta(0)$. It shows that the increase in value λ will also increase the value of $\theta(0)$. It can be observed that the solution is not unique for a certain critical value λ in both figures. There is no solution to the equation obtained when $\lambda < \lambda_c$ due to separation of the boundary layer from the surface.

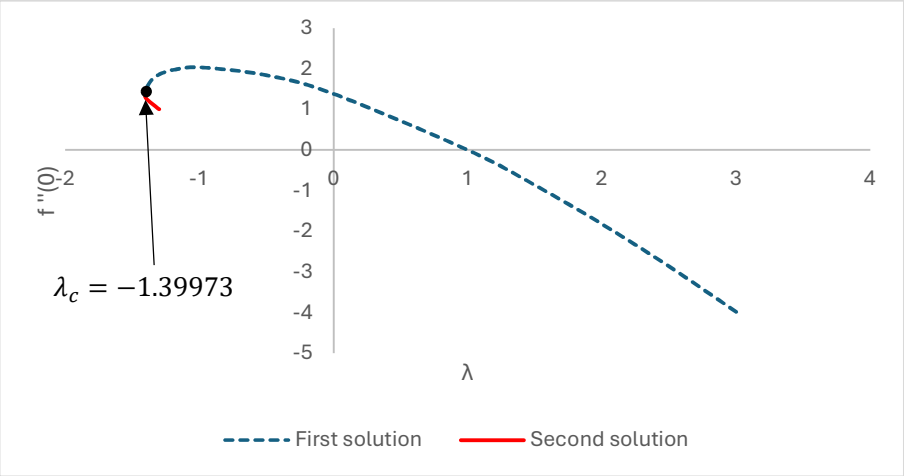


Figure 3.5: The skin friction coefficient, $f''(0)$ versus λ when $\beta = 0.1$

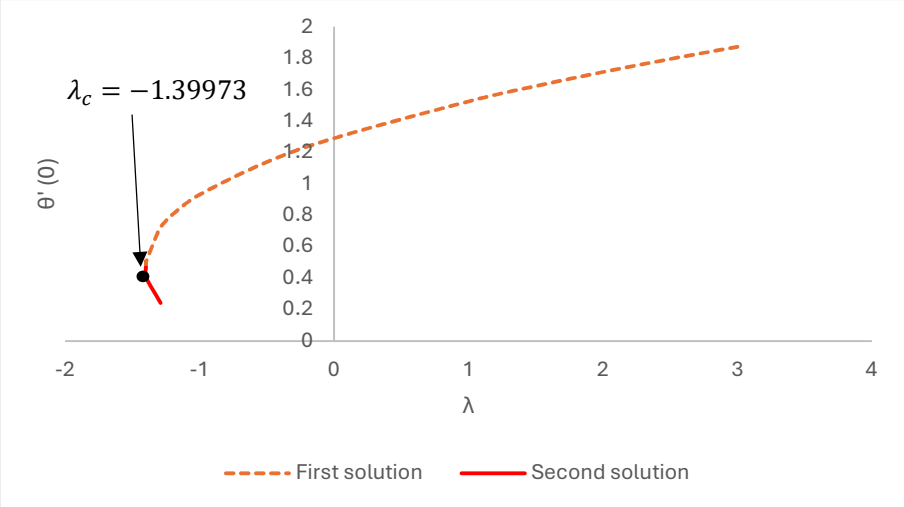


Figure 3.6: The local Nusselt number, $\theta'(0)$ versus λ when $\beta = 0.1$

3.2 Effect of Hartree pressure gradient, β on velocity profile, temperature profile, skin friction coefficient and local Nusselt number.

The velocity profile and temperature profile graphs are shown in Figures 3.7 and Figure 3.8 with $M = 0.2, s = 1$, $\lambda = -1.30$ with $\beta = 0.4, 3.0, 7.5$. The velocity profile in Figure 3.7 illustrates how increases in β will result in higher flow velocity profiles close to the wedge's surface. When β increases, the thickness of the boundary layer decreases, indicating a greater velocity gradient near the surface. However, in Figure 3.8, the graph shows that the local Nusselt number decreases as the β increases. It demonstrates that β has a significantly lower impact on temperature profiles than velocity profiles.

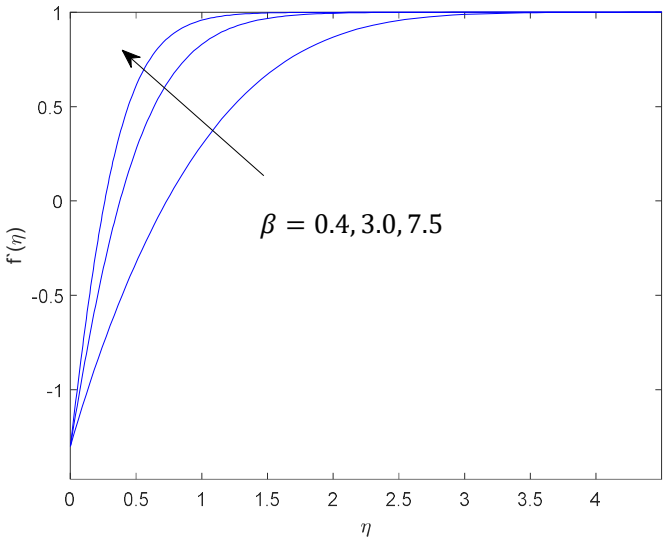


Figure 3.7: The velocity profile for different value of β, when $\lambda = -1.30$

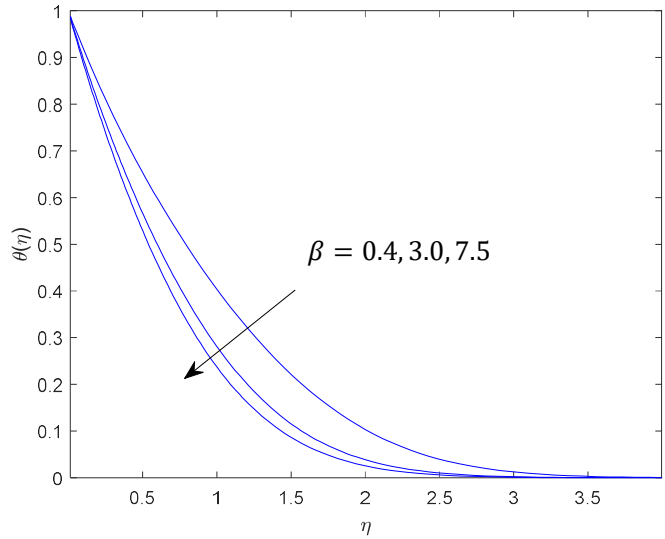


Figure 3.8: The temperature profile for different value of β, when $\lambda = -1.30$

4. Conclusion

This study investigated magnetohydrodynamic (MHD) flow over shrinking wedges. The use of similarity reduction variables transforms the nonlinear equations, known as PDEs, into ODEs. The equations are then numerically resolved using MATLAB's integrated BVP4C function. In addition to the effects of local Nusselt number and skin friction, temperature and velocity profiles impacted by different parameter values especially in shrinking parameter, λ and Hartree pressure gradient, β were observed. The results were also compared to earlier research to determine the validity of BVP4C's results. The results are validated through their lose agreement with previous studies.

5. Acknowledgments

The authors would like to thank the Faculty of Computer Science and Mathematics, Universiti Teknologi Mara, Dr. Siti Hidayah Binti Muhad Saleh as supervisor for the support.

6. References

[1] Ibrahim, W., & Tulu, A. (2019). Magnetohydrodynamic (MHD) Boundary Layer Flow Past a Wedge with Heat Transfer and Viscous Effects of Nanofluid Embedded in Porous Media. *Mathematical Problems in Engineering*, *2019*(1), 1-12. https://doi.org/10.1155/2019/4507852

[2] Karthik, S., Iranian, D., Alhazmi, H., Khan, I., & Singh, A. (2024). Heat transfers due to electromagnetic radiation of MHD Powell-Eyring fluid with mass diffusion via Lie symmetry scaling. *Case Studies in Thermal Engineering*, *58*(1), 1-16. https://doi.org/10.1016/j.csite.2024.104348

[3] Rasheed, H. (2017). MHD Boundary Layer Flow of Nanofluid over a Continuously Moving Stretching Surface. *Applied and Computational Mathematics*, *6*(6), 265. https://doi.org/10.11648/j.acm.20170606.15

[4] Shahzad, A., Habiba, B., Nadeema, M., Kamranb, M., Ahma, H., Atifd, M., & Ahmad, S. (2021). Numerical Analysis of Flow and Heat Transfer in a Thin Film Along an Unsteady Stretchig Cylinder. *Thermal Science*, *25* (2), 441–448. https://doi.org/10.2298/TSCI21S2441S

[5] Alisman, N., Fazahiyah, N., & Shab, M. (2023). Numerical Analysis on MHD Stagnation-Point Flow Towards a Stretching Sheet using bvp4c Method with Slip Effect. *Sciences and Technology*, *3*(2), 16–024. https://doi.org/10.30880/ekst.2023.03.02.003

[6] Akiruddin, N. H., & Aman, F. (2023). Solving MHD Flow over a Shrinking Wedges Using Shooting Technique. *Sciences and Technology*, *3*(1), 057–063. https://doi.org/10.30880/ekst.2023.03.01.007

[7] Awaludin, I. S., Ishak, A., & Pop, I. (2018). On the Stability of MHD Boundary Layer Flow over a Stretching/Shrinking Wedge. *Scientific Reports*, *8*(1), 1-8. https://doi.org/10.1038/s41598-018-31777-9

[8] Parvin, S., Isa, S. S. P. M., Al-Duais, F. S., Hussain, S. M., Jamshed, W., Safdar, R., & Eid, M. R. (2022). The flow, thermal and mass properties of Soret-Dufour model of magnetized

Maxwell nanofluid flow over a shrinkage inclined surface. *PLos One*, *17*(4), 1-25. https://doi.org/10.1371/journal.pone.0267148

[9] Nayak, M. K., Abdul Hakeem, A. K., & Makinde, O. D. (2019). Time varying chemically reactive magneto-hydrodynamic non-linear falkner-skan flow over a permeable stretching/shrinking wedge: Buongiorno model. *Journal of Nanofluids*, *8*(3), 467–656. https://doi.org/10.1166/jon.2019.1616

[10] Amar, N., & Kishan, N. (2021). The influence of radiation on MHD boundary layer flow past a nano fluid wedge embedded in porous media. *Partial Differential Equations in Applied Mathematics*, *4*(1), 1-8. https://doi.org/10.1016/j.padiff.2021.100082

[11] Bejan, A. & Kraus, A. D. (2003). *Heat transfer handbook*. John Wiley. https://pdfcoffee.com/heat-transfer-handbook-by-adrian-bejan-amp-allan-d-kraus-pdf-free.html

[12] Hussain, M., Ali, A., Yao, S. W., Ghaffar, A., & Inc, M. (2022). Numerical investigation of chemically dissipated mixed convective flow. *Case Studies in Thermal Engineering*, *31,101809*. https://doi.org/10.1016/j.csite.2022.101809

[13] Srinivasacharya, D., Mendu, U., & Venumadhav, K. (2015). MHD Boundary Layer Flow of a Nanofluid Past a Wedge. *Procedia Engineering*, *127*(1), 1064–1070. https://doi.org/10.1016/j.proeng.2015.11.463

Modelling the Common Edible Oil Price based on the US Dollar Index

Aina Arinah Riduwan[1,2] and Yusrina Andu[3,*]

[1]*Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan, Malaysia.*

[2]*Institute For Clinical Research, Block B4, National Institutes of Health (NIH), No.1, Jalan Setia Murni U13/52, Seksyen U13,40170 Shah Alam, Selangor Darul Ehsan Malaysia*

[3]*Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Kuala Pilah, Negeri Sembilan, Malaysia.*

[*]Corresponding author: yusrinaandu@uitm.edu.my

Abstract

Edible oil is a type of lipid which can be used to prepare food and is derived from plants, animals, or microorganisms. It is a heart-healthy type of cooking oil that can handle high heat while still maintaining its nutritional value. The study of edible oil prices is critical because it shows the demand and supply, which can affect the economy of a country. Meanwhile, the US Dollar Index is also important as it is a global currency, and any changes that occur can affect other commodity prices. Therefore, this study will investigate the relationship between edible oil prices and the US Dollar Index. Additionally, the best model will also be determined. Using a sample of 229 observations, this study explored the use of multiple linear regression in analysing the standard edible oil price based on the US Dollar Index. The result indicated that the best models for the US Dollar Index are the prices of coconut, peanut, rapeseed and sunflower oil. In conclusion, this model is valuable for traders, producers, consumers, and policymakers to forecast price trends and make informed decisions based on currency movements and market conditions.

Keywords*:* Edible oil, Modelling, US Dollar

Introduction

Edible oils play a crucial role in the human diet as they provide essential dietary fats that contribute to overall health, supporting growth and brain development. These oils, also known as cooking oils, are derived from plants, animals, or microorganisms and are used in food preparation [1]. Popular edible oils consumed worldwide include avocado, safflower, grapeseed, sesame, and extra virgin olive oil. These oils are desired due to their ability to handle high heat during cooking, making them heart-healthy options. The consumption of edible vegetable oil has increased over the years, as reported by the National Bureau of Statistics, which states that China consumed 39.78 million tonnes of edible vegetable oil between 2018 and 2019 [2]. This highlights the significant role of edible vegetable oil in daily nutrition, particularly in Asian diets.

In fiscal year 2022, vegetable oil sales in India amounted to approximately $2.6 billion, reflecting a significant increase from the previous year's sales of over $1.8 billion. It was

anticipated that the sales value would continue to rise, reaching around $3.2 billion [3]. According to [4], the prices of sunflower oil, soybean oil, palm oil, and others are found in significant global edible oil trading centres. Edible oils, like coconut oil, peanut oil, olive oil, and palm oil, are typically traded globally, and their prices are often quoted in US Dollars. Therefore, the strength of the US Dollar plays a crucial role in determining the cost of these commodities on the international market.

Generally, researchers and analysts have employed various methods to predict the price of edible oils. These methods included the use of statistical models, supply and demand analysis, and the application of machine learning algorithms. A study using the autoregressive conditional heteroskedasticity (GARCH) method [5] discovered a negative correlation between crude oil prices and the value of the US Dollar, indicating their tendency to move together. The GARCH specification, although not based on any economic theory, offered a good approximation to the heteroskedasticity commonly seen in financial time-series data. Another study in [6] focuses on model-based simulations, as it is appropriate for trade policy analysis to forecast the economic implications of yet-to-be-implemented policy changes. This considers domestic production, price, income and government revenue of edible oil as the variables.

On the other hand, poor farm workers can significantly benefit from price forecasting by using it to assist them in deciding what to grow and when to sell [7]. In their study, artificial neural networks (ANN) and Auto Regressive Integrated Moving Average (ARIMA) were used to create forecasting models for the prices of three types of edible oils. Next, a study from [8], which forecasted the crude palm oil price, mentioned three forecasting techniques of multivariate time series. It was evaluated by comparing the evaluation metrics of MSE, RMSE, MAE and MAPE, and the lowest error value of the evaluation metrics determines the best model. In [9], a multiple log-linear regression model is used to determine the relationship between sunflower edible oil demand, sunflower and palm oil prices, and per capita income. Another study by [10] on oil's adulterant analysis uses simple linear regression (SLR), multiple linear regression (MLR), and partial least squares regression (PLSR) to predict lard in palm olein oil. Thus, this shows that modelling is also common in edible oil studies.

Linear regression is a popular statistical method which can model the relationship between explanatory and response variables in various studies. By analysing historical prices of edible oils, researchers could uncover the connection between different types of edible oils and their prices. However, fewer studies have been carried out which apply linear regression modelling to edible oils data with the US Dollar Index. Therefore, this study aims to determine the best model that explains the relationship between these two variables.

Methodology

Data description

The research was conducted using secondary data obtained from the Kaggle website, consisting of 7 variables and 229 observations, covering the period from July 2003 to July 2022. The details and descriptions of each variable can be found in Table 1.

Table 1: Description of Variables

No.	Variable Name	Variable Type	Variable	Description of price data
1	US Dollar Index	Interval	Dependent	monthly and yearly
2	Coconut oil price	Interval	Independent	monthly
3	Olive oil price	Interval	Independent	monthly
4	Palm kernel oil price	Interval	Independent	monthly
5	Peanut oil price	Interval	Independent	monthly
6	Rapeseed oil price	Interval	Independent	monthly
7	Sunflower oil price	Interval	Independent	monthly

Theoretical Framework

The research findings showed that the prices of coconut, olive, palm kernel, peanut, rapeseed, and sunflower oil had an impact on the US Dollar Index. This index was considered the dependent variable, while the oil prices served as independent variables. Figure 1 illustrates the relationship and the variable labels.

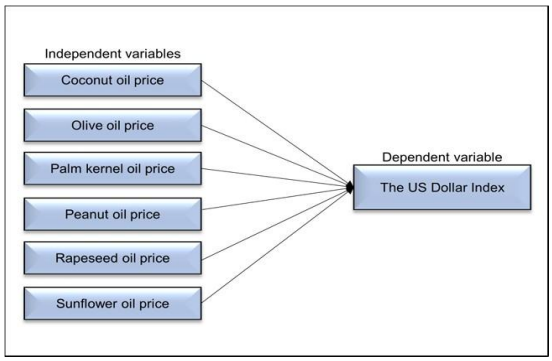


Figure 1: Theoretical Framework

Multiple Linear Regression

Multiple linear regression (MLR) is used to model the linear relationship between the dependent variable and the six independent variables. This relationship can be described as in Eq. 1:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \quad (1)$$

where Y_i is US Dollar Index, X_1 = coconut oil price, X_2 = olive oil price, X_3 = palm kernel oil price, X_4 = peanut oil price, X_5 = rapeseed oil price, X_6 = sunflower oil price and ε = error term.

If the *p*-value is less than 0.05, the null hypothesis is rejected, and it can be concluded that there is a significant difference in the model. To evaluate the model, it is necessary to make sure that all the assumptions of the regression model are fulfilled. These four assumptions for

MLR conditions that need to be fulfilled are that the error variable must be normally distributed, the error variance must be constant, the error terms must be independent, and no multicollinearity between the variables exists.

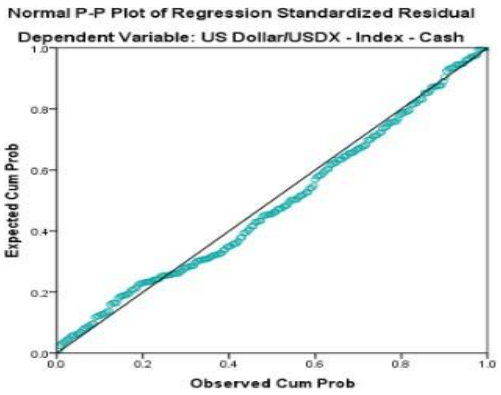
Stepwise regression is a technique in MLR used to identify the most important predictor variables and remove less significant ones from a large group. It is known to be computationally effective as it helps build a more straightforward and more effective model, particularly when there are many potential predictors [11]. Here, this method will be used as it can present the model simply and more understandably due to its capability of selecting a small selection of essential predictor variables. To test the significance of the regression model, Analysis of Variance (ANOVA) is used. This is to ensure that the model is valid before being used for forecasting or making inferences. The efficiency of the model predictive value is estimated by using root mean squared error (RMSE). Finally, the proportion of the variance in the dependent variable that is explained by the independent variables in the model is measured by using the coefficient of multiple determination, R^2.

Results

Model adequacy checking

Four assumptions need to be tested between the US Dollar Index, coconut, olive, palm kernel, peanut, rapeseed and sunflower oil. The regression model was used since all assumptions are fulfilled. The first assumption is fulfilled as it shows that the error variable is normally distributed (Figure 2a), since the points align closely with the diagonal line. The homoscedasticity assumption is also met as the pattern in Figure 2b shows a random dispersion of residuals around zero, with no discernible pattern or trend.

a)



b)

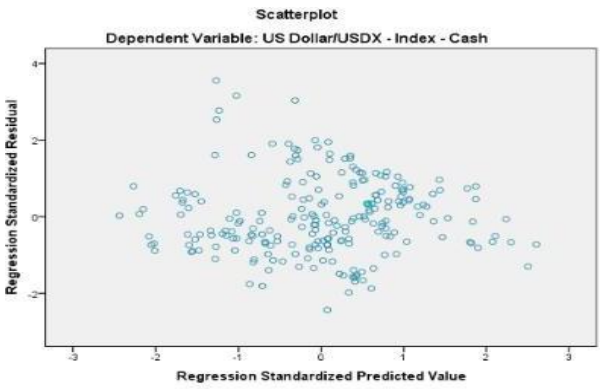


Figure 2: a) Normal P-P Plot of Regression Standardised Residuals b) Scatter Plot of Dependent Variable

For the Durbin-Watson (DW) statistic, if the value of the statistic is between 1.5 and 2.5, it can be assumed that the error terms are independent. The value calculated is 1.531, indicating

that they are independent since it is within the range. Thus, satisfying this assumption. Finally, there is no multicollinearity detected between the independent variables because all the Variance Inflation Factors are less than 10 and the tolerance scores are greater than 0.1. This is present in Table 2.

Table 2: Multicollinearity Test for Independent Variable

Independent Variable	Collinearity Statistic	
	Tolerance	VIF
Coconut oil price	0.475	2.107
Peanut oil price	0.301	3.326
Rapeseed oil price	0.116	8.619
Sunflower oil price	0.114	8.798

Model validity checking

The validity of the model is tested using the F-test regression and is presented using ANOVA in Table 3.

Table 3: ANOVA table result

	Sum of Squares	df	Mean Square	F	Sig.
Regression	6.219.063	4	1554.766	41.307	0.000
Residual	8431.277	224	37.640		
Total	14650.340	228			

It can be observed that the p-value is less than 0.05, therefore indicating the validity of the model. There is a statistically significant difference between the means of the variable in this model. Meanwhile, the RMSE value is 8431.277, obtained as the lowest value, thus making it a preferable model compared to others. Meanwhile, the R2 value obtained between the independent and dependent variables is 0.424. This can be described as, 42.4% of the US Dollar Index can be explained by the price of coconut, peanut, rapeseed and sunflower oil, while other factors explain the remaining 57.6%.

The next step is testing the significance of each of the parameters after validating the model. This part is essential because it determines the importance of the independent variables in estimating the dependent variable. All variables are significant at a p-value less than 0.05. This shows that the price of coconut, peanut, rapeseed and sunflower oil is significantly affecting the US Dollar Index. The coefficient of the significant regression variable is given as MLR in Equation (2):

$$\hat{Y} = 243.734 + 0.011 CoconutOilPrice + 0.009 PeanutOilPrice + 9.484 RapeseedOilPrice - 36.029 SunflowerOilPrice \quad (2)$$

The coconut, peanut and rapeseed oil prices all have a positive impact on the US Dollar index, with the rapeseed oil having a higher value. This may indicate that it is the primary contributor

to the US Dollar index. Meanwhile, the price of sunflower oil harms the US Dollar index. This suggests that the increase in sunflower oil price may cause a reduction in the US Dollar index, which may be due to the cost adjustments in the market.

Conclusion

Thus, by using MLR, this study has successfully shown how different types of edible oils can influence the overall market prices. A stronger US Dollar index will cause both coconut oil and rapeseed oil prices to become more expensive, while sunflower oil prices will become cheaper. It can be concluded that global exporters can adjust their strategies based on the US Dollar fluctuations, as a weaker US Dollar might increase their sales. This model can also assist stakeholders or policy makers in predicting the future price movements of edible oil and the US Dollar index.

References

[1] Čabarkapa-Pirković, A., Živković, L., Dekanski, D., Topalović, D., & Spremo-Potparević, B. (2021). Olive leaf, DNA damage and chelation therapy. In *Olives and Olive Oil in Health and Disease Prevention* (pp. 457-469). Academic Press.

[2] Xia, Q., Du, Z., Lin, D., Huo, L., Qin, L., Wang, W., ... An, Y. (2021). Review on contaminants in edible oil and analytical technologies. Oil Crop Science, 6(1), 23-27.

[3] Statista. (2022). Statista. India: Vegetable oil sales value from 2012 to 2020. Retrieved from https://www.statista.com/statistics/1365662/ india-vegetable-oil-sales-value/

[4] Yu, T.-H. E., Bessler, D. A., & Fuller, S. W. (2006). Cointegration and causality analysis of world vegetable oil and crude oil prices.

[5] Le, D. (2017). Relationship between Crude Oil Prices and the US Dollar Exchange Rates: Constant or Time-varying?. *Journal of Applied Finance and Banking*, *7*(5), 103.

[6] Indhushree, A., & Shivakumar, K. M. (2020). Dynamics of palm oil import on prices, income and trade of Indian edible oil sector. *Indian Journal of Agricultural Economics*, *75*(4), 578-586.

[7] Singh, A. (2021). Comparison of artificial neural networks and statistical methods for forecasting prices of different edible oils in indian markets. *International Research Journal of Modernization in Engineering Technology and Science*, *3*, 1044-1050.

[8] Kanchymalay, K., Salim, N., Sukprasert, A., Krishnan, R., & Hashim, U. R. A. (2017, August). Multivariate time series forecasting of crude palm oil price using machine learning techniques. In *IOP Conference series: materials science and engineering* (Vol. 226, No. 1, p.

012117). IOP Publishing.

[9] Mgeni, C. P., & Mpenda, Z. T. (2021). Can Sub-Saharan Africa become food self-sufficient? Analyzing the market demand for sunflower edible oil in Tanzania. *Tanzania Journal of Agricultural Sciences*, *20*(1), 42-53.

[10] Topcu, Y., Turhan, B., & Uzundumlu, A. S. (2010). Analysis of factors affecting Turkish sunflower oil consumer behavior: the case of Erzurum province. *Italian Journal of Food Science/Rivista Italiana di Scienza degli Alimenti*, *22*(2).

[11] Fox, J. (2019). *Regression diagnostics: An introduction*. Sage publications.

A COMPARATIVE STUDY OF FIRST ORDER DIFFERENTIAL EQUATION BY USING BUTCHER'S FIFTH ORDER METHOD AND DORMAND PRINCE METHOD

Norlin Anas[1], Rahmah Shahril[2*], Maznah Banu Mohamed Habiboo Raman[3], Rasidah Buang[4], Rosha Mohamed[5]

[1,2,3,5]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan

[4]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Cawangan Melaka, Kampus Jasin, Melaka

Corresponding author: [*]rahma949@uitm.edu.my

Abstract

Nonlinear differential equations present considerable challenges when tackled with fundamental approaches due to their inherent complexity. Numerical methods like Euler, Taylor, and Runge-Kutta provide effective alternatives for solving these equations. The Butcher's fifth-order method and Dormand Prince method are part of the Runge-Kutta family that have been used to address higher-order differential equations and have been applied in science and mathematics fields. Both methods are known for their high accuracy in solving such problems, as they can produce better approximation to the exact solutions. However, there are limited studies on solving the first-order Bernoulli and Riccati differential equations using these methods. Therefore, this study attempts to solve both differential equations using the Butcher's fifth-order and Dormand Prince methods with a fixed step size of h = 0.1. Both methods were analysed by comparing their absolute errors against the exact solution. The results showed that both Butcher's and Dormand Prince methods are capable of solving first-order Bernoulli and Riccati differential equations. However, in terms of accuracy, the findings highlight that the Butcher's fifth-order method is better at solving problems 1 and 3, while the Dormand Prince method is suitable for solving problems 2 and 4, demonstrating more accurate approximations due to smaller errors. It proved that both methods are accurate and suitable for solving these problems. However, the accuracy of both methods depends on the specific problems. Future research could investigate their application to higher-order nonlinear differential equations with different step sizes. The accuracy and efficiency of both methods could be further evaluated by comparing relative errors, computational work, or execution time.

Keywords*:* Differential equation, Bernoulli and Riccati problems, Error analysis, Butcher's fifth order method, Dormand Prince method

Introduction

Many real-world situations in science and engineering can be modelled using differential equations. Differential equations are particularly crucial to scientists and engineers, as they serve as essential tools for the mathematical modelling of any circumstances requiring a rate of change [1]. A differential equation is known as an equation that involves an independent variable, a dependent variable, and the derivative of the dependent variable with respect to the independent variable. These differential equations can be classified into two categories: ordinary differential equations (ODEs) and partial differential equations (PDEs) [2]. An ODE is a mathematical equation that expresses one or more functions of an independent variable along with its derivatives. ODEs can be categorised into linear and nonlinear equations. There are various types of nonlinear differential equations such as second-order Abel's differential equation, Chrystal's equation, and Lagrange's

equation. However, this study will focus on first-order nonlinear differential equations, specifically Bernoulli and Riccati differential equations.

Bernoulli's differential equation, introduced by Jacob Bernoulli, can be transformed from a nonlinear equation into a linear equation by substitution, where $z = y^{1-n}$, where $n \neq 0$ and $n \neq 1$ [3]. This equation is still used in real-life applications which are covered in mathematics, engineering and physics [4]. Applications include the analysis of free vibration of Euler-Bernoulli beams modelled by spatial-fractional differential equations [5], identifying the properties of q-differential equations of higher order using q-Bernoulli polynomials [6], and identifying daily traffic flow patterns based on the nonlinear grey Bernoulli model [7]. Meanwhile, Riccati differential equation is named after Jacopo Francesco Riccati. This equation has also been applied to various fields including mathematics, physics and engineering, such as optimal control, robust stability, and finding numerical solutions of Riccati differential equations using Euler and Runge-Kutta methods [8]. This equation can be solved analytically by various numerical methods such as Adomian decomposition method [9], He's variational iteration method (VIM), the homotopy perturbation method (HPM) and the iterated He's HPM [10], and a piecewise VIM [11].

In this study, Butcher's fifth order and Dormand Prince methods are used to solve the first order Bernoulli and Riccati differential equations and the results obtained will be compared to the exact solutions by calculating their errors based on absolute error.

Methodology

Two methods Butcher's fifth order and Dormand Prince methods will be represented in this section for solving first order Bernoulli and Riccati differential equations.

Butcher's fifth order method

Butcher's fifth order method is an extension of the Runge-Kutta fourth order (RK4) method, which consists of six function evaluations of intermediate slopes in obtaining the average weighted function values at different points. The article by Hossain et al. (2017) was referred to obtain the solution of the differential equations. The mathematical formula of Butcher's method is demonstrated as follows:

$$y_{n+1} = y_n + \frac{h}{90}\left(7k_1 + 32k_3 + 12k_4 + 32k_5 + 7k_6\right) \text{ for } n = 0,1,2,... \tag{1}$$

where:

$$k_1 = f(x_n, y_n)$$
$$k_2 = f(x_n + \frac{1}{4}h, y_n + \frac{1}{4}k_1 h)$$
$$k_3 = f(x_n + \frac{1}{4}h, y_n + \frac{1}{8}k_1 h + \frac{1}{8}k_2 h)$$
$$k_4 = f(x_n + \frac{1}{2}h, y_n - \frac{1}{2}k_2 h + k_3 h)$$
$$k_5 = f(x_n + \frac{3}{4}h, y_n + \frac{3}{16}k_1 h + \frac{9}{16}k_4 h)$$
$$k_6 = f(x_n + h, y_n - \frac{3}{7}k_1 h + \frac{2}{7}k_2 h + \frac{12}{7}k_3 h - \frac{12}{7}k_4 h + \frac{8}{7}k_5 h) \tag{2}$$

Dormand Prince method

The Dormand Prince method is one of the Runge-Kutta (RK) methods that has been widely applied in real-life applications, especially in the fields of science and chemistry, due to its high accuracy. This method involves a

seven-stage function evaluations of intermediate slopes at different points to obtain the estimated solutions. However, this method has its own property known as First Same As Last (FSAL) property, only six function evaluations need to be calculated [13]. The mathematical formula of Dormand Prince method was referred from Keong and Hamzah (2018) and is presented below:

$$y_{n+1} = y_n + \frac{35}{384}k_1 + \frac{500}{1113}k_3 + \frac{125}{192}k_4 - \frac{2187}{6784}k_5 + \frac{11}{86}k_6 \text{ for } n = 0,1,2,... \tag{3}$$

where:

$$k_1 = hf(x_n, y_n)$$
$$k_2 = hf(x_n + \frac{1}{5}h, y_n + \frac{1}{5}k_1)$$
$$k_3 = hf(x_n + \frac{3}{10}h, y_n + \frac{3}{40}k_1 + \frac{9}{40}k_2)$$
$$k_4 = hf(x_n + \frac{4}{5}h, y_n + \frac{44}{45}k_1 - \frac{56}{15}k_2 + \frac{32}{9}k_3)$$
$$k_5 = hf(x_n + \frac{8}{9}h, y_n + \frac{19372}{6561}k_1 - \frac{25360}{2187}k_2 + \frac{64448}{6561}k_3 - \frac{212}{729}k_4)$$
$$k_6 = hf(x_n + h, y_n + \frac{9017}{3168}k_1 - \frac{355}{33}k_2 + \frac{46732}{5247}k_3 + \frac{49}{176}k_4 - \frac{5103}{18656}k_5)$$
$$k_7 = hf(x_n + h, y_n + \frac{35}{384}k_1 + \frac{500}{1113}k_3 + \frac{125}{192}k_4 - \frac{2187}{6784}k_5 + \frac{11}{84}k_6) \tag{4}$$

Error Analysis

The results obtained from the proposed methods will be compared with the exact solutions based on absolute error. The purpose of this comparison is to determine which method is more accurate and suitable for solving the first order Bernoulli and Riccati differential equations. The accuracy of the results can be assessed by observing the smaller errors produced by the proposed methods. The formula for absolute error is presented below:

$$E = \left| y - \bar{y} \right| \tag{5}$$

where:

E = Absolute error
y = Actual value
$\bar{y}$ = Predicted value

Illustrative Problem

The effectiveness of the proposed methods for solving first order Bernoulli and Riccati equations will be presented in this section. All the computations are performed using Maple 18 software, with the step size fixed at $h = 0.1$.

Problem 1

Consider the Bernoulli differential equation taken from [14].

$$y' = x^3 y^3 - xy \tag{6}$$

with the initial conditions $y(0)=1$, and the exact solution is given by $y(x) = \frac{1}{1+x^2}$.

Table 3.1 Solutions of problem 1 with a step size of h = 0.1 and $x \in [0,1]$

x	Exact solution	Butcher's Fifth Order Method	Dormand Prince Method	Errors of Butcher's Method	Errors of Dormand Prince Method
0.0	1.0000000000	1.0000000000	1.0000000000	0	0
0.1	0.9900990099	0.9950371888	0.9950371905	4.9382×10^{-3}	4.9382×10^{-3}
0.2	0.9615384615	0.9805806735	0.9805806762	1.9042×10^{-2}	1.9042×10^{-2}
0.3	0.9174311927	0.9578262830	0.9578262858	4.0395×10^{-2}	4.0395×10^{-2}
0.4	0.8620689655	0.9284766890	0.9284766914	6.6408×10^{-2}	6.6408×10^{-2}
0.5	0.8000000000	0.8944271893	0.8944271914	9.4427×10^{-2}	9.4427×10^{-2}
0.6	0.7352941176	0.8574929238	0.8574929260	1.2220×10^{-1}	1.2220×10^{-1}
0.7	0.6711409396	0.8192319180	0.8192319206	1.4809×10^{-1}	1.4809×10^{-1}
0.8	0.6097560976	0.7808688061	0.7808688094	1.7111×10^{-1}	1.7111×10^{-1}
0.9	0.5524861878	0.7432941421	0.7432941461	1.9081×10^{-1}	1.9081×10^{-1}
1.0	0.5000000000	0.7071067763	0.7071067810	2.0711×10^{-1}	2.0711×10^{-1}

Problem 2

Consider the Bernoulli differential equation taken from [15].

$$y' = 2xy + 2x^3y^2 \tag{7}$$

with the initial condition $y(0)$=1, and the exact solution is given by $y(x) = \frac{1}{1-x^2}$.

Table 3.2 Solutions of problem 2 with a step size of h = 0.1 and $x \in [0, 0.9]$

x	Exact solution	Butcher's Fifth Order Method	Dormand Prince Method	Errors of Butcher's Method	Errors of Dormand Prince Method
0.0	1.0000000000	1.0000000000	1.0000000000	0	0
0.1	1.0101010100	1.0101010190	1.0101010100	9.0000×10^{-9}	0
0.2	1.0416666670	1.0416666890	1.0416666650	2.2000×10^{-8}	2.0000×10^{-9}
0.3	1.0989010990	1.0989011470	1.0989010940	4.8000×10^{-8}	5.0000×10^{-9}
0.4	1.1904761900	1.1904763010	1.1904761790	1.1100×10^{-7}	1.1000×10^{-8}
0.5	1.3333333330	1.3333336150	1.3333332950	2.8200×10^{-7}	3.8000×10^{-8}
0.6	1.5625000000	1.5625008090	1.5624998480	8.0900×10^{-7}	1.5200×10^{-7}
0.7	1.9607843140	1.9607867560	1.9607834970	2.4420×10^{-6}	8.1700×10^{-7}
0.8	2.7777777780	2.7777757860	2.7777733360	1.9920×10^{-6}	4.4420×10^{-6}
0.9	5.2631578950	5.2616611750	5.2635061440	1.4967×10^{-3}	3.4825×10^{-4}

Problem 3

Consider the Riccati differential equation taken from [16].

$$y' = -\sin x + \left(\cos^2 x\right)y - \left(\cos x\right)y^2 \tag{8}$$

with the initial condition $y(0)$=1, and the exact solution is given by $y(x) = \cos x$.

Table 3.3 Solutions of problem 3 with a step size of $h = 0.1$ and $x \in [0,1]$

x	Exact solution	Butcher's Fifth Order Method	Dormand Prince Method	Errors of Butcher's Method	Errors of Dormand Prince Method
0.0	1.0000000000	1.0000000000	1.0000000000	0	0
0.1	0.9950041653	0.9950041644	0.9950041660	9.0000×10^{-10}	7.0000×10^{-10}
0.2	0.9800665778	0.9800665764	0.9800665792	1.4000×10^{-9}	1.4000×10^{-9}
0.3	0.9553364891	0.9553364873	0.9553364910	1.8000×10^{-9}	1.9000×10^{-9}
0.4	0.9210609940	0.9210609920	0.9210609963	2.0000×10^{-9}	2.3000×10^{-9}
0.5	0.8775825619	0.8775825600	0.8775825646	1.9000×10^{-9}	2.7000×10^{-9}
0.6	0.8253356149	0.8253356132	0.8253356178	1.7000×10^{-9}	2.9000×10^{-9}
0.7	0.7648421873	0.7648421858	0.7648421902	1.5000×10^{-9}	2.9000×10^{-9}
0.8	0.6967067093	0.6967067081	0.6967067122	1.2000×10^{-9}	2.9000×10^{-9}
0.9	0.6216099683	0.6216099672	0.6216099710	1.1000×10^{-9}	2.7000×10^{-9}
1.0	0.5403023059	0.5403023049	0.5403023084	1.0000×10^{-9}	2.5000×10^{-9}

Problem 4

Consider the Riccati differential equation taken from [17].

$$y' = 1 + y^2 \tag{9}$$

with the initial condition $y(0)=0$, and the exact solution is given by $y(x) = \tan x$.

Table 3.4 Solutions of problem 4 with a step size of $h = 0.1$ and $x \in [0,1]$

x	Exact solution	Butcher's Fifth Order Method	Dormand Prince Method	Errors of Butcher's Method	Errors of Dormand Prince Method
0.0	0.0000000000	0.0000000000	0.0000000000	0	0
0.1	0.1003346721	0.1003346723	0.1003346720	2.0000×10^{-10}	1.0000×10^{-10}
0.2	0.2027100355	0.2027100366	0.2027100354	1.1000×10^{-9}	1.0000×10^{-10}
0.3	0.3093362496	0.3093362523	0.3093362494	2.7000×10^{-9}	2.0000×10^{-10}
0.4	0.4227932187	0.4227932240	0.4227932183	5.3000×10^{-9}	4.0000×10^{-10}
0.5	0.5463024898	0.5463024987	0.5463024887	8.9000×10^{-9}	1.1000×10^{-9}
0.6	0.6841368083	0.6841368221	0.6841368053	1.3800×10^{-8}	3.0000×10^{-9}
0.7	0.8422883805	0.8422884002	0.8422883722	1.9700×10^{-8}	8.3000×10^{-9}
0.8	1.0296385570	1.0296385820	1.0296385330	2.5000×10^{-8}	2.4000×10^{-8}
0.9	1.2601582180	1.2601582360	1.2601581510	1.8000×10^{-8}	6.7000×10^{-8}
1.0	1.5574077250	1.5574076800	1.5574075250	4.5000×10^{-8}	2.0000×10^{-7}

The results from the proposed methods are displayed in Tables 3.1 – 3.4. The numerical solutions and errors are calculated with a fixed step size of 0.1 by using Maple 18 software. The results are compared to the exact solutions. Butcher's fifth order method gives a better approximation to the exact solution for problems 1 and 3, while Dormand Prince method demonstrates a close approximation to the exact solution for problems 2 and 4. In terms of accuracy, Butcher's fifth order method is more accurate and suitable for solving problems 1 and 3. For problems 2 and 4, Dormand Prince method is more accurate and suitable, presenting a better approximation with smaller errors produced.

Conclusion

In this paper, Butcher's fifth order and Dormand Prince methods have been applied to solve first order Bernoulli and Riccati differential equations. It can be concluded that both methods are capable of solving all the problems. However, in terms of accuracy, the accuracy of both methods varies depending on the specific problem.

References

[1] Mengesha, L. M., & Denekew, S. A. (2020). Revised methods for solving nonlinear second order differential equations. *Journal of Applied & Computational Mathematics*, *9*(4), 1-4. doi: 10.37421/jacm.2020.9.456.

[2] Devi, S., & Jakhar, M. (2018). An introduction to differential equations. *International Journal of Statistics and Applied Mathematics*, *3*(1), 115–117. https://www.researchgate.net/publication/359578950

[3] Azarian, M. K. (2012). On Bernoulli's differential equation. *Applied Mathematical Sciences*, *6*(137), 6799–6803. https://www.m-hikari.com/ams/ams-2012/ams-137-140-2012/azarianAMS137-140-2012-1.pdf

[4] Arbab, S. Y., & Altoum, S. H. (2020), q -Bernoulli's equation. *American Journal of Applied Sciences*, *17*(1), 214–217. doi: 10.3844/ajassp.2020.214.217.

[5] Jafari, A., & Aftabi Sani, A. (2024). Free vibration analysis of Euler-Bernoulli beams modeled by spatial-fractional differential equation. *Results in Engineering*, *24*. doi: 10.1016/j.rineng.2024.102972.

[6] Ryoo, C. S., & Kang, J. Y. (2022). Properties of q-differential equations of higher order and visualization of fractal using q-Bernoulli polynomials. *Fractal and Fractional*, *6*(6). doi: 10.3390/fractalfract6060296.

[7] Yang, L., Su, B., Xie, N., & Wei, B. (2024). An IDE-based nonlinear grey Bernoulli model and applications to daily traffic flow pattern identification. *Expert Systems with Applications*, *246*. doi: 10.1016/j.eswa.2024.123172.

[8] Batiha, B. (2015). A new efficient method for solving quadratic Riccati differential equation. *International Journal of Applied Mathematical Research*, *4*(1), 24. doi: 10.14419/ijamr.v4i1.4113.

[9] El-Tawil, M. A., Bahnasawi, A. A., & Abdel-Naby, A. (2024). Solving Riccati differential equation using Adomian's decomposition method. *Applied Mathematics Computation*, *157*(2), 503–514. doi: 10.1016/J.AMC.2003.08.049.

[10] Abbasbandy, S. (2006). Homotopy perturbation method for quadratic Riccati differential equation and comparison with Adomian's decomposition method. *Applied Mathematics and Computation*, *172*(1), 485–490. doi: 10.1016/J.AMC.2005.02.014.

[11] Geng, F. (2010). A modified variational iteration method for solving Riccati differential equations. *Computers and Mathematics with Applications*, *60*(7), 1868–1872. doi: 10.1016/j.camwa.2010.07.017.

[12] Hossain, Md. J., Alam, Md. S., & Hossain, Md. B. (2017a). A study on numerical solutions of second order initial value problems (IVP) for ordinary differential equations with fourth order and Butcher's fifth order Runge-Kutta methods. *American Journal of Computational and Applied Mathematics*, *2017*(5), 129–137. doi: 10.5923/j.ajcam.20170705.02.

[13] Keong, A. T., & Hamzah, N. S. A. (2018). Solving ordinary differential equations by the Dormand Prince method. *A Letter on Applications of Mathematics and Statistics.*

[14] Abu-Amr, S., Almajbri, T., & Mohammed, A. (2024). Bernoulli differential equations solution using Adomian decomposition method by Matlab. *AlQalam Journal of Medical and Applied Sciences*, 794–798. doi: 10.54361/ajmas.247348.

[15] Nasr Al Din (2020). Numerical study for solving Bernoulli differential equations by using Runge-Kutta Method and "Newton's Interpolation and Aitken's Method." *International Journal of Scientific and Innovative Mathematical Research*, *8*(5). doi: 10.20431/2347-3142.0805002.

[16] Masjed-Jamei, M., & Salehi Shayegan, A. H. (2017). A numerical method for solving Ricatti differential equations. *Iranian Journal of Mathematical Sciences and Informatics*, *12*(2), 51–71. doi: 10.7508/ijmsi.2017.2.004.

[17] Nasr Al Din (2020). Numerical study for solving quadratic Riccati differential equations. *Middle-East Journal of Scientific Research*, *28*(4), 348–356. doi: 10.5829/idosi.mejsr.2020.348.356.

The Price of Research Octane Number 97 (RON97) in Malaysia: A Univariate Time Series Approach

Nurul Syakinah Binti Mohd Zaki[1], Noreha Mohamed Yusof[1,*]
[1]*Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, 70300, Seremban, Negeri Sembilan, Malaysia.*
[*]noreh144@uitm.edu.my

Abstract

Modelling RON97 fuel prices can be challenging as it sometimes can be particularly difficult to select a suitable method to achieve accurate predictions. Different countries exhibit varying patterns of petrol (gasoline) usage, adding complexity to the forecasting process. The study emphasized the importance of selecting models that balance accuracy with simplicity while addressing data characteristics like trends. This research explored the effectiveness of various time series methods in predicting RON97 fuel prices by using Double Exponential Smoothing (DES) and Holt's Method. This paper tried to identify the most accurate and practical model for forecasting. Data that was being used was monthly data of RON97 price from January 2020 until January 2024. The error measurements which were Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) showed that the lowest value in deciding the best method within the two methods was Holt's method. Thus, with an MSE value of 0.0002977 and MAPE value of 0.1653 which was the smallest value. The result is significant for government of Malaysia and also to all Malaysian for energy security.

Keywords*:* RON97, univariate models, double exponential smoothing, holts' method

Introduction

The study of gasoline prices has long been a focal point for both theoretical and applied economists for numerous reasons [1]. The pricing dynamics of fuel exhibit considerable heterogeneity, both temporally and spatially, as observed across different geographical regions and over time [2]. This heterogeneity fosters a localized competition within the gasoline sector, where gas stations primarily regard nearby counterparts as their main competitors [3].

In Malaysia, there are three types of gasoline which are Diesel, RON95 and RON97. The findings indicated that the RON95 is the cheapest and economic compared to RON97 [4]. All three have a different type fuel. As for the RON97 industrial fuel is a bit different as this fuel has being widely used in automotive and industrial applications. The price of RON97 fuel, is a premium-grade gasoline that widely use. Even though it can be a bit different, this fuel is a critical determinant in economic activity, transportation cost and consumer spending patterns.

Ringgit Malaysia fluctuations more or less give impact to the fuel price particularly for non-subsidized fuels such as RON97, because its cost is directly connected to global crude oil prices, which are traded in US dollars. When the ringgit falls against the dollar, the cost of importing or acquiring crude oil and processed products rises, resulting in higher domestic fuel costs [5]. Given that global crude oil is primarily traded in US dollars, fluctuations in the Ringgit

Malaysia (MYR) against the USD can significantly impact the cost of imported crude oil for Malaysia. For non-subsidized fuels like RON97, which are directly linked to global crude prices, these exchange rate variations are likely to influence domestic RON97 prices. Conversely, if the Ringgit strengthens, it can help to stabilize or even cut fuel prices, making it less expensive to import these goods. This link demonstrates the sensitivity of fuel costs to currency exchange rates and other economic factors that influence the Ringgit's value.

Additionally, this study with facilitate people of learning new knowledge. This research surely increased people knowledge in gasoline and fuel terms. Through the research of past, findings and others, there is a lot of knowledge that can be obtained and use in daily life. Lastly, this study with facilitate the researcher on the scientific knowledge and opportunities for innovation in investigating RON97. Researcher also find a scientific contribution in understanding a new theory in the industry. This study addresses the need to determine the best forecasting model for monthly fuel price RON97 between Double Exponential Smoothing (DES) and Holt's Method model.

Methodology

Data Source

This study utilized secondary data that was collected from [6] website. Data were used in this research project was the monthly mean concentration price of RON97 in Malaysia starting from 4th January 2020 to 4th January 2024. Data was was obtained from the open data portal website.

Double Exponential Smoothing

This method is also known as Brown's Method. It is helpful for series that show signs of a linear trend. Below are the model's equations:

$$S_t = \alpha y_t + (1-\alpha)S_{t-1} \tag{1}$$

$$S_t^{'} = \alpha S_t + (1-\alpha)S_{t-1} \tag{2}$$

$$a_t = 2S_t - S_t^{'} \tag{3}$$

$$b_t = \frac{\alpha}{1-\alpha}\left(S_t - S_t^{'}\right) \tag{4}$$

$$F_{T+m} = a_T + b_T \times m \tag{5}$$

Holt's Method

The use of a single smoothing constant in Brown's method can lead to estimate linear trend values that are susceptible to random variations. Holt's two-parameter method was another commonly used data handling technique to get around this issue. This method was also chosen as it was suitable for use and was commonly used for analysis based on past research. Therefore, this study decides to apply this method to analyze the data.

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{6}$$

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1} \quad (7)$$

$$F_{T+m} = S_T + T_T \times m \quad (8)$$

Comparing Holt's Method and Double Exponential Smoothing

The evaluation of the model's forecast performance was conducted using Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) values to decide the best model between Holt's Method and the Double Exponential Smoothing Method. To assess model forecast performance using the one-step-ahead procedure, the equation for MSE was as follows,

$$MSE = \frac{\sum_{t=1}^{n} e_{t+1}^2}{n} \quad (9)$$

While for the MAPE, the equations was as follows:

$$MAPE = \sum_{t=1}^{n} \frac{(e_t)/(y_t) \times 100}{n} \quad (10)$$

where n denotes effective data points and |(et)/(yt) = 100| is defined as the absolute percentage error obtained from the fitted values for a certain forecasting method.

Analysis

Preliminary Analysis

The line chart of price of RON97 and trend values has been constructed in Figure 1.

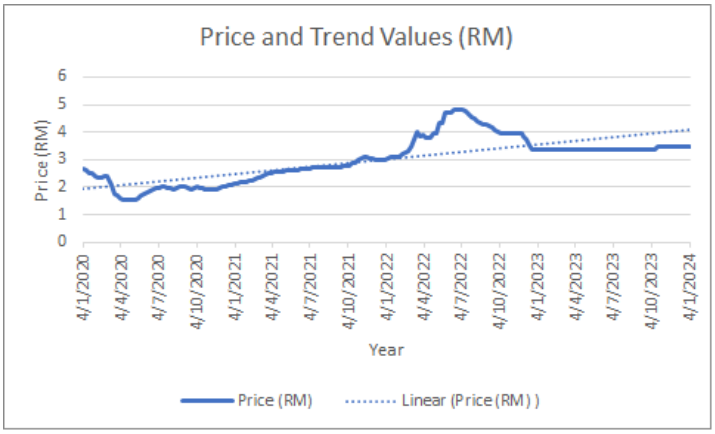


Figure 1: Price and Trend Values of RON97.

Line chart in Figure 1 shows that the trend line suggests a gradual upward trend in prices, indicating a general increase over the observed period. The line chart shows clear trend

component with a gradual upward movement in prices over time, as the line chart have no cyclical, no irregular and no seasonal trend. However, the actual price data shows significant fluctuations, particularly around June 2022 until August 2022, where prices reach a peak before declining sharply.

From late 2022 onward, the price stabilizes at a lower level, showing minimal fluctuations and suggesting that it may have reached a new equilibrium. This stability could indicate a balance in supply and demand or reduced market pressures that previously drove the price up.

Double Exponential Smoothing Method

The trend line plot of fitted value versus actual value of RON97 price described in Figure 2 using Double Exponential Smoothing Method.

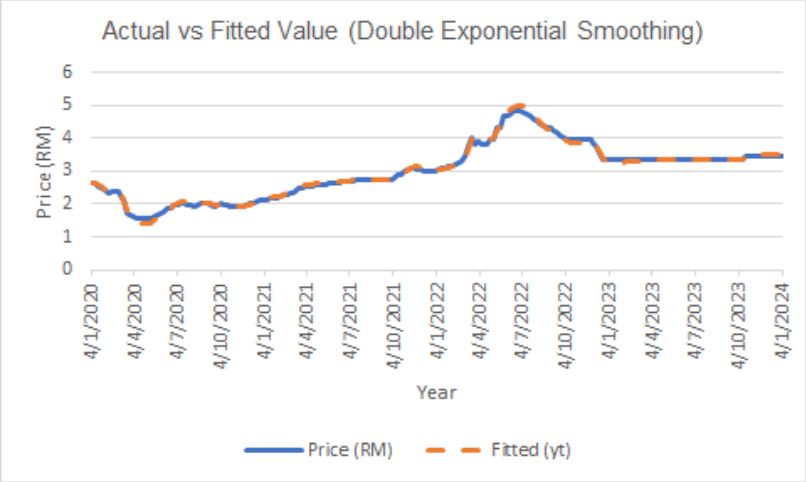


Figure 2: Actual and Fitted of RON97 prices using Double Exponential Smoothing.

The Double Exponential Smoothing (DES) model demonstrates strong accuracy in capturing the historical data of RON97 prices, as seen in the close alignment between the actual data and the fitted values. The model effectively adapts to the significant trends in the data, including the upward price trajectory starting in 2021 and peaking in early 2022, followed by a sharp decline later in 2022. After this decline, the model reflects a stabilization in prices from 2023 onward, which is projected to continue into 2024. The close match between the actual and fitted lines suggests that DES is well-suited for analysing and forecasting time series data with trends but without seasonality.

Holt's Method

The trend line plot of Fitted value versus Actual value using Holts Method of RON97 prices has shown in Figure 3.

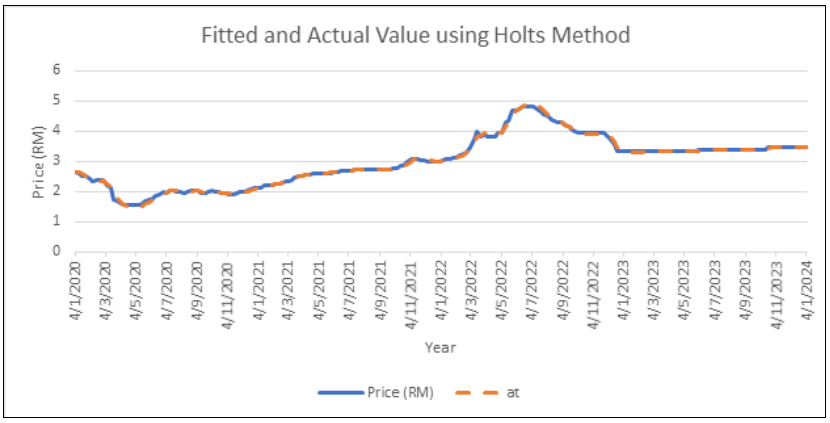


Figure 3: Actual and Fitted of RON97 prices using Holt's Method

The analysis using Holt's method shows a strong alignment between the fitted values and the actual RON97 prices, indicating that the method effectively captures both the level and trend components of the data. The model successfully tracks the upward price trend from 2021 to the peak in early 2022, as well as the subsequent sharp decline later that year. From 2023 onward, the model reflects a period of stabilization in prices, with the forecast extending this stability into 2024.

Comparing Holt's Method and Double Exponential Smoothing

The MSE and MAPE values for both Double Exponential Smoothing (DES) and Holt's Method are different, but all the values are nearly to zero, which is expected as they represent the same forecasting technique which is under the exponential smoothing technique. Using evaluation part (February 2023 until January 2024), to compare between the two methods.

Table 1: Summary of MSE and MAPE value of Double Exponential Smoothing and Holt's method in Evaluation part

Method	MSE	MAPE
Double Exponential Smoothing	0.0007076	0.4935
Holt's Method	0.0002977	0.1653

To conclude the summary of the table shows that for the value of Mean Square Error (MSE), Holt's method has the smallest value compared to the Double Exponential Smoothing method value. Holt's method also has the smallest value for MAPE compared to the Double Exponential Smoothing method. In conclusion, Holt's method has been chosen as the best method as it has the smallest value for MSE and MAPE compared to the Double Exponential Smoothing Method.

Conclusion

It was crucial to comprehend the price of the RON97 trend in Malaysia to know its fluctuations in growth and decline. The pattern encompassed a trend line and its components. Both upward

and downward movements represent elements of the trend. From early January 2020 until May 2020 the trend declined as the start of the pandemic of COVID-19 happened. The trend pattern starts to grow rapidly until it starts to drop again. From the middle of June 2022 until August 2022, it grew rapidly, Malaysia was dealing with its economic challenges, including managing the fiscal deficit and ensuring economic stability. These factors, combined with global influences, contributed to the fluctuations in fuel prices also Malaysia, like many other countries, was experiencing high inflationary pressures. The rising cost of commodities, including fuel, contributed to the overall increase in prices. The government had to manage these pressures while maintaining fuel subsidies [7].

The Double Exponential Smoothing model has significant alignment between real and fitted values, which successfully captures historical data trends. DES approach is best suited for datasets that exhibit trends but lack seasonality. Its capacity to handle these qualities makes it a dependable option for short- to medium-term forecasting. The low error metrics, such as MSE, RMSE, and MAPE, suggest that the model closely approximates the observed data while minimizing forecasting errors. Regardless of its virtues, DES assumes a linear trend, which may limit its performance in datasets with more complex patterns. [8] stated that using Double Exponential Smoothing can improve the Artificial Neural Network machine learning to exchange rate forecasting.

Holt's Method is highly accurate, with somewhat better performance metrics than DES. Holt's method builds on DES by adding a second smoothing parameter for trend estimation, making it significantly more advanced. Its higher performance indicators, such as lower MSE and MAPE, indicate that it captures the dataset's trend component more effectively than DES. [6], stated that Holt's Method is suitable to use within the seasonal trend and can improve the forecast accuracy in analyse the current dataset.

Based on the analysis, it clearly shown that Holt's method is the better model compared to Double Exponential Smoothing to visualize the price of RON97 in Malaysia.

References

[1] A. S. Bergantino, C. Capozza, and M. Intini, "Empirical investigation of retail fuel pricing: The impact of spatial interaction, competition and territorial factors," *Energy Economics*, vol. 90, p. 104876, 2020.

[2] A. Bergeaud and J. Raimbault, "An empirical analysis of the spatial variability of fuel prices in the United States," *Transportation Research Part A: Policy and Practice*, vol. 132, pp. 131–143, 2020.

[3] A. F. Colladon, G. Verdoliva, L. Segneri, and A. G. Vitali, "Analyzing gasoline prices in five Italian cities: Insights from social network analysis," *Cities*, vol. 150, p. 105075, 2024.

[4 A. K. Rashid, M. R. A. Mansor, A. Racovitza, and R. Chiriac, "Combustion characteristics of various octane rating fuels for automotive thermal engines efficiency requirements," *Energy Procedia*, vol. 157, pp. 763–772, 2019.

[5] Bank Negara Malaysia, *Economic and Monetary Review 2023*. Kuala Lumpur, Malaysia: Bank Negara Malaysia, 2023. [Online]. Available: https://www.bnm.gov.my/documents/20124/12141961/emr2023_en_book.pdf.

[6] Government of Malaysia, "Price of petroleum & diesel," Jul. 2024. [Online]. Available: https://www.kpdn.gov.my. [Accessed: Jul. 11, 2024].

[7] SME Bank Malaysia, "Malaysia under high oil prices and inflationary environment," 2023. [Online]. Available: https://www.smebank.com.my/documents/d/guest/magazine_malaysia_under_high_oil_prices inflationary_environment. [Accessed: Jan. 14, 2025].

[8] S. Muladi, S. A. Siregar, and A. P. Wibawa, "Double exponential-smoothing neural network for foreign exchange rate forecasting," in *Proc. 2nd East Indonesia Conf. Comput. Inf. Technol. (EIConCIT)*, 2018, pp. 118–122. doi: 10.1109/EIConCIT.2018.8878591.

OPTIMISING WATER SUPPLY AND DEMAND ALLOCATION USING DYNAMIC PROGRAMMING: ADDRESSING DISTRIBUTION INEFFICIENCIES

Aiman Syazwan Ahmad[1,*], Rossidah Wan Abdul Aziz[1]

[1]College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA, Cawangan Negeri Sembilan,
Kampus Seremban, 70300 Seremban Negeri Sembilan,
Malaysia
Email: rossidah@uitm.edu.my

Abstract

Water resource allocation is a critical issue for domestic, agricultural, and industrial sectors, particularly in regions facing water scarcity. This study aims to address the problem of inefficiency in water distribution among competing sectors, leading to unmet demands and resource wastage. The main objective of this study is to find an optimal allocation of water resources using a Dynamic Programming (DP) model to enable the available water supply to be distributed efficiently to meet the demand of each sector. A structured approach was employed to achieve this aim, beginning with a broad literature review of the prevailing strategies for allocation and how DP has been applied. Secondary data on the supply and demand of water resources from credible sources over the past five years were used. The DP model was then used to decompose this complex problem into more manageable subproblems to define optimal allocation strategies. The results obtained indicate that the proposed DP model effectively optimizes water distribution in every sector. Furthermore, data analysis results reflect that water demand is highly responsive to available water. This can be inferred from the fact that the maximum water demand is more sensitive to changes in the available water supply than the minimum demand, hence requiring adaptive management. In conclusion, this study provides valuable insights into the dynamics of water allocation and offers a robust framework that could enhance efficient distribution. Therefore, the adoption of the DP model in water management strategies by policymakers is highly recommended since it was complemented by real-time data monitoring to adapt to changing water availability and demands.

Keywords: Water resource allocation, Dynamic Programming model, Bellman equation, supply and demand

1. Introduction

The term "water resources" refers to the amount of water that is available for ecosystems and people. Humans have access to and use water for a variety of purposes, including domestic use, industry, agriculture, recreation, and environmental conservation. Another definition states that water is an essential natural resource required to raise the standard of living for urban residents and promote economic growth at the local level. The availability of water is often uncertain, and it can be considered a limited resources; therefore, distribution control must be done in an orderly manner [1]. Water allocation is a way of distributing limited resources among multiple consumers who require it at the same time. It is crucial to allocate time effectively when distributing water, as this can ensure its security, reduce its wastage and maximize its community benefits [2].

Even so, there are several major factors influence the allocation of this water, including ongoing population growth, which increases demand for water supplies. Furthermore, water pollution caused by community and industrial activities prevents the even distribution of clean water. One way to allocate

resources effectively is to assess the quantity and quality of water that is available. Water must first be allocated to meet basic needs before it can be used for other purposes. Therefore, an equitable allocation of water resources across regions is essential for reducing poverty sustainably as well as promoting economic growth, particularly in areas where water demands are rising rapidly [3]. However, there are still significant differences in water accessibility.

Despite technological advancements, the challenge of balancing water resource allocation with the growing human population and industry remains unresolved. Increasing demand due to population growth leads to water shortages and causing difficulties in distribution, particularly in highly populated and industrialized regions. This imbalance forces communities to compete for quality water resources, threatening long-term sustainability and future water security. Water scarcity is a critical issue that hampers social and economic progress, exacerbated by continuous urbanization and industrial activities [4]. Addressing this issue requires optimizing water allocation while considering demand fluctuations and response uncertainties. Therefore, this study focuses on Dynamic Programming (DP) model to maximize the allocation of water supply for each sector. This study also aims to analyze the total water supply allocated according to demand for each sector by using linear regression model.

2. Literature Review

Water resources management is always indispensable in terms of the allocation of the available amount among different consumers and industries with the purpose of ensuring efficiency for long-term use. [5] evaluated the hydrological effects of climate change on reservoir operation by using the hybrid model which Stochastic Dynamic Programming (SDP) model and the Soil and Water Assessment Tool (SWAT). In light of climate change, this hybrid model optimizes reservoir operations to enhance agricultural water supply and distribution.

Integrating various goals could help preserve sustainable water consumption while also attending to different priorities of stakeholders [6]. The ability of algorithms to reproduce optimal equilibrium solutions under a range of risk circumstances helps the watershed managers in process of making decisions. Moreover, the dynamic scheduling approach was employed to take decisions with respect to the current circumstances for proper management of water resources when there is drought or unpredictable climatic changes at certain seasons [3]. The Multi-Objective Dynamic Programming model that used in that research enable to ensures the availability of water resources in the future while meeting the current sector of water demands by optimizing water utilization efficiency.

Richard Bellman first presented DP in 1950, and it has been widely used in a variety of fields, including computer science, mathematics, operations research, economics, and engineering. This method is an effective approach to address complex problems by breaking the problem down into smaller problems that are solved once. But in order to prevent recomputing, their solutions are stored and used in recursion.

Xu, [7] involved special focus on agricultural consumption optimization during water management in linking reservoirs. They solve the difficulties related to multidimensional optimization problems in reservoirs by using the improved Decomposition and Dynamic Programming Aggregation model. They found that water shortage can be minimize through optimization of water supply and reduction in system energy consumption. Other than that, [8] was focused on the optimization of water resource management in the river basin. The SDP model and water value provide a comprehensive approach to water allocation methods in an area where competing demands and growing water shortages coexist. Their study looks at water consumption as well as pricing evaluation in the domestic, agricultural and industrial sectors.

In conclusion, the applicability and efficiency of these models in solving complicated allocation problems are analyzed. This review aims to provide a theoretical basis and identify knowledge gaps that justify the adoption of dynamic programming in this study.

3. Methodology

3.1. Data Collection

This study utilised secondary data from [6]. The data collected for Major River Basins (MRB) are water transport loss ratio, available water resources for each period, and minimum and maximum water demands by region in the MRB for all three sectors.

3.2. DP model

In the modelling of the water resource allocation, we divide the work involved into two phases that are developing conceptual framework based on [9] as present in Figure 1, and modeling the DP model for water resources allocation.

Forward Dynamic Programming is used to solve the main problems which would start from the first stage and move forward stage by stage to find the optimal allocation for each stage. Initially, this model is solved for the first period by obtaining the total allocation of water supply in that period and then employing forward recursion by adding the total allocation from period 1 with total allocation for period 2. This process has been repeated until periods 3 and 4.

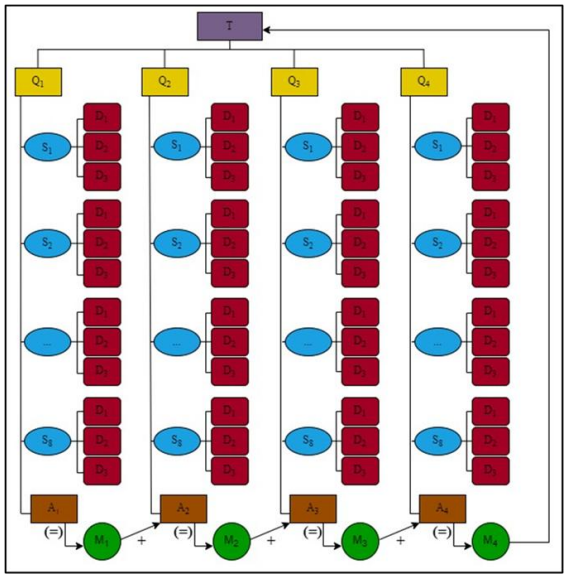


Figure 1: DP Structure

3.3. Notations of the model

Indices

t: Index of period t of the regional water resources allocation where $t = 1, 2, 3, 4$, $t = 1$: Spring; $t = 2$: Summer; $t = 3$: Autumn; $t = 4$: Winter.

i: Index of subarea or region i in the river basin where $i = 1, 2, \dots, 8$, $i = 1$: Chengdu; $i = 2$: Zigong; $i = 3$: Neijiang; $i = 4$: Leshan; $i = 5$: Meishan; $i = 6$: Yibin; $i = 7$: Ya'an; $i = 8$: Aba.

j: Index of water sector in the subarea i where $j = 1, 2, 3$, $j = 1$: Domestic sector; $j = 2$: Agriculture sector; $j = 3$: Industry sector.

Parameters

Q_t: Allocation of water supply from the river basin in period t.
$S_{t,i}$: Allocation of water supply for subarea i in period t.
$D_{t,i,j}^{min}$: Minimum demand of water supply in period t for subarea i for sector j.
$D_{t,i,j}^{max}$: Maximum demand of water supply in period t for subarea i for sector j.
A_t: Actual allocation of water supply in period t.
M_t: Total amount of water supply allocated for use in period t.
$\rho_{t,i}$: Water-transfer loss ratio from river basin in period t for subarea i.
$\alpha_{t,i}^{e}$: Water allocated to the sustainable development of the entire basin ecological environment of subarea i in period t
s_t: Available water supply in the river basin at the beginning of each period t.
s_{t+1}: Available water supply in the river basin for the next period t.
s_t^{max}: Maximum volume of stored water in period t.

Decision variable

$D_{t,i,j}$: Demand of water supply in period t for subarea i for sector j.
T: Total allocation of water supply from river basin in a year.

Based on DP structure, this study has derived the objective function of the model. Objective function for water allocation of DP is to find the maximum allocation of water supply for three sectors in eight subareas in a year. After employed the recursive function, the general Bellman's equation or the objective function for the model can be formulated as follows:

$$Maximise\ T = \begin{cases} \sum_{t=1}^{4}\sum_{i=1}^{8}\sum_{j=1}^{3} D_{t,i,j} & ,t = 1,2,3,4, i = 1,2,\dots,8\ and\ j = 1,2,3 \\ \sum_{t=1}^{4}\sum_{i=1}^{8}\sum_{j=1}^{3} D_{t,i,j} + M_{t-1} & ,t = 2,3,4, i = 1,2,\dots,8\ and\ j = 1,2,3 \end{cases} \quad (1)$$

3.4. Formulation of Constraints

(1) Limitation on the water supplies that are available

The total amount of water supply that allocated to each sector in each subarea in one period, $D_{t,i,j}$ cannot be more than the total amount of water supply that allocated from the river basin to that period, Q_t.

$$0 < \sum_{t=1}^{4}\sum_{i=1}^{8}\sum_{j=1}^{3} D_{t,i,j} \leq Q_t \quad (2)$$

(2) Restriction of water consumption for each subarea per period

The summation between total amount of water supply that allocated to each sector, $D_{t,i,j}$ and the water allocated to the ecological environment in that subarea, $\alpha_{t,i}^{e}$ cannot be greater than the amount allocated to that subarea, $S_{t,i}$.

$$\sum_{t=1}^{4}\sum_{i=1}^{8}\sum_{j=1}^{3} D_{t,i,j} + \alpha_{t,i}^{e} \leq S_{t,i} \quad (3)$$

(3) Water demand limitations

The amount of water allocated to each sector should fall between the minimum water demand, $D_{t,i,j}^{min}$

and the maximum water demand, $D_{t,i,j}^{max}$.

$$D_{t,i,j}^{min} \leq D_{t,i,j} \cdot (1 - \rho_{t,i}) \leq D_{t,i,j}^{max} \quad (4)$$

(4) Equation describing state transfer

The available water supply for the next period is equal to available water supply at the beginning of each period, s_t minus with the total amount of water supply that allocated to each sector in each subarea, $D_{t,i,j}$.

$$s_{t+1} = s_t - \sum_{t=1}^{4} \sum_{i=1}^{8} \sum_{j=1}^{3} D_{t,i,j} \quad (5)$$

(5) The non-negative restriction must also be met by each individual variable.

$$D_{t,i,j} \geq 0 \quad , \quad T \geq 0 \quad (6)$$

4. Result and Discussions

4.1. Total Water Allocation for each Period

The MRB has variable water availability throughout the year due to rainfall and other natural factors. Water management is thus necessary to supply the demand for the domestic, agriculture and industry sectors without harming the environment. Table 1 shows the result for total amount of water allocated during spring, summer, autumn and winter.

Table 1: Results for total allocation of water supply for each period

Period	Water availability ($\mathbf{10^4 m^3}$)	Total water allocation ($\mathbf{10^4 m^3}$)
Spring	60931.86	55770.49
Summer	290940.52	267839.30
Autumn	750859.10	653132.80
Winter	125009.10	117848.00

During spring season, the water availability in spring is relatively low at 60,931.86 $\times 10^4 m^3$ with most of it 55,770.49 $\times 10^4 m^3$ allocated according to the water demand from each sector. This indicates that there was leaving little surplus at this period due to a high reliance on the water resources that were available. Similarly in summer, water availability increases significantly to 290,940.52 $\times 10^4 m^3$ and water allocations increase correspondingly to 267,839.30 $\times 10^4 m^3$. The comparatively greater surplus in comparison to spring indicates that there may be more water resources available to satisfy peak demands especially for irrigation which frequently occur during this season.

In autumn, this season stands out with the highest water availability which 750,859.10 $\times 10^4 m^3$ and highest total water allocation which 653,132.80 $\times 10^4 m^3$ due to increased rainfall and storage capacity. Even with the large allocation, there is still a substantial surplus which may be the result of decreased demand or inability to meet the water demand in particular sectors. For winter, this season sees the lowest water availability 125,009.10 $\times 10^4 m^3$ but maintains relatively high allocation levels 117,848 $\times 10^4 m^3$ leaving only a small surplus of water resource. This pattern can be the result of a sustained demand due to domestic and industrial consumption with reduced natural inflow due to dry weather. Overall, the water allocation strategy seems well-judged as that would meet seasonal demands but at the same time leaving a consistent surplus to help meet the long-term water sustainability.

4.2. Total Water Allocation for Eight Subareas during All Period

The results of the water allocation throughout the regions and sectors during all seasons where spring, summer, autumn and winter have to be examined in order to ensure a fair and efficient distribution that meets all water demands. The results of this study will be discussed in the context of the comparison of allocated water resources to minimum required water demands of the domestic, agricultural and industrial sectors of each of the eight subareas, assessing the efficiency and sufficiency of the allocation model. The objective is to identify the locations where allocations have not been able to meet the demands and need further optimization. The results of water allocation for all subareas are as shown in Table 2.

Table 2: Results for water allocation in eight subareas during different period ($10^4 m^3$)

Period	**Sector**	Chengdu	Zigong	Neijiang	Leshan	Meishan	Yibin	Yaan	Aba
Spring	Dom	5795.39	768.58	1017.10	835.28	653.72	2376.46	421.92	264.67
	Agr	12597.51	1998.31	2028.99	1994.23	2918.41	5660.78	1149.82	480.45
	Ind	5795.39	886.14	920.74	971.02	870.53	4548.21	563.65	253.22
Summer	Dom	24423.80	3959.74	6560.95	5985.54	3673.17	7833.08	2626.23	2709.34
	Agr	53090.33	10295.32	15794.14	14290.49	16398.08	18658.57	7156.99	5015.93
	Ind	24423.80	4565.34	7319.85	6958.19	4891.31	14991.40	3508.32	2709.34
Autumn	Dom	82806.70	7909.75	6250.16	17774.22	10426.08	8937.94	6679.97	2437.67
	Agr	179997.90	20565.33	14673.71	42435.95	46544.97	21290.38	18204.20	4512.98
	Ind	82806.70	9119.47	6745.30	20662.54	13883.70	17105.94	8923.62	2437.67
Winter	Dom	11415.02	1721.56	2085.74	2523.87	1642.90	3695.86	1434.58	812.20
	Agr	24812.99	4476.05	5124.95	6025.73	7334.42	8803.65	3909.49	1503.68
	Ind	11415.02	1984.85	2201.71	2933.99	2187.75	7073.37	1916.42	812.20

The water allocation strategy is generally well-distributed across seasons but several allocations fail to meet the minimum demand with the autumn season showing the worst performance in meeting water demand compared to other seasons. Several subareas were generally going through a serious shortage especially in Chengdu, Zigong, Neijiang and Meishan which highlighted as critical areas for improvement in allocation strategies. Agriculture is the most affected sector and it experiences a shortfall every other season, which will impact decreased crop yields and economic losses to farmers with the potential to cause harm to local food security. Although the domestic sector often receives allocations near the minimum demand, shortages still occur in every season. This can affect the access of household to a sufficient amount of water supply for drinking, cooking and hygiene purposes during peak periods of demand. Shortage of water is comparatively more severe during the autumn season compared to spring, summer and winter. However, the water allocation in all seasons needs better management and conservation of infrastructure to ensure all sectors get adequate water.

5. Conclusion and Recommendations

In summary, this study provided an essential insight in the management of water resource in terms of maximizing distribution to the industrial, agricultural, as well as domestic sectors. This indicates that the study successfully achieved its intended goals and providing valuable insights to water resource allocation. The DP method provided a useful framework for allocating water resources in the most efficient way possible, besides giving an in-depth examination into the dynamics of water supply and demand in modelling water resources allocation under different seasonal constraint. The findings validate the effectiveness of the DP method as a robust tool for optimizing water allocation especially in scenarios with fluctuating supply and demand across seasons.

Acknowledgment

The authors would like to take this opportunity to express their heartfelt gratitude to the staff and lecturers of KPPIM at Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Campus Seremban, who provided invaluable support and solutions for the successful completion of this Final Year Project.

References

[1] W. Li *et al.*, "Chance-constrained dynamic programming for multiple water resources allocation management associated with risk-aversion analysis: A case study of Beijing, China," *Water (Switzerland)*, vol. 9, no. 8, pp. 1–16, Aug. 2017, doi: 10.3390/w9080596.

[2] L. Ji, P. Sun, Q. Ma, N. Jiang, G. H. Huang, and Y. L. Xie, "Inexact two-stage stochastic programming for water resources allocation under considering demand uncertainties and response-A case study of Tianjin, China," *Water (Switzerland)*, vol. 9, no. 6, pp. 1–14, Jun. 2017, doi: 10.3390/w9060414.

[3] J. Feng, "Optimal allocation of regional water resources based on multi-objective dynamic equilibrium strategy," *Appl Math Model*, vol. 90, pp. 1183–1203, Feb. 2021, doi: 10.1016/j.apm.2020.10.027.

[4] C. Ren, Y. Wang, L. Yu, H. Zhang, and Z. Xie, "An improved bi-level programming model for water resources allocation under multiple uncertainties," *J Environ Manage*, vol. 345, pp. 1–12, Nov. 2023, doi: 10.1016/j.jenvman.2023.118882.

[5] W. T. Ayele, D. Y. Tenagashaw, A. Z. Belew, and T. G. Andualem, "Optimal Water Allocation Under Climate Change, Based on Stochastic Dynamic Programming Model Approach in Ribb Reservoir, Amhara, Ethiopia," *Water Conservation Science and Engineering*, vol. 7, no. 1, pp. 33–44, Mar. 2022, doi: 10.1007/s41101-021-00125-w.

[6] M. Yuan, X. Chen, H. Ren, X. Zhou, and Z. Yan, "Dynamic multi-period sustainable water resources optimal allocation strategies: A case study of China," *Comput Ind Eng*, vol. 186, pp. 1–18, Dec. 2023, doi: 10.1016/j.cie.2023.109713.

[7] Z. Xu, Z. Gong, H. Cheng, and J. Cheng, "Optimal water allocation integrated with water supply, replenishment, and spill in the in-series reservoir based on an improved decomposition and dynamic programming aggregation method," *Journal of Hydroinformatics*, vol. 25, no. 3, pp. 1–15, May. 2023, doi: 10.2166/hydro.2023.208.

[8] C. Davidsen, S. J. Pereira-Cardenal, S. Liu, X. Mo, D. Rosbjerg, and P. Bauer-Gottwein, "Using Stochastic Dynamic Programming to Support Water Resources Management in the Ziya River Basin, China," *J Water Resour Plan Manag*, vol. 141, no. 7, pp. 1–12, Jul. 2015, doi: 10.1061/(asce)wr.1943-5452.0000482.

[9] R. W. A. Aziz, A. Shuib, A. H. Nawawi, and N. M. Tawil, "Mathematical Model for Budget Planning and Execution," *Journal of Industrial and Intelligent Information*, vol. 3, no. 2, pp. 1–6, June. 2015, doi: 10.12720/jiii.3.2.91-96.

OPTIMIZING VEHICLE PARKING CAPACITY IN NARROW SPACE: AN ANALYSIS OF THE BINGLE CONCEPT

Luqman Hakim Muhamad Kamal[1] and Najir Tokachil[2,*]

[1,2]College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan.
*najir2992@uitm.edu.my

Abstract

One of the essential urban issues that cities face worldwide is parking. The number of vehicles is increasing mainly due to the trend of people immigrating to the cities. However, this problem is mainly caused by the limited parking space area. Thus, to figure out this issue, the parking lot of the FSKM, Universiti Teknologi MARA(UiTM) Seremban 3, was selected to be part of this study. The Bingle Concept was implemented to optimize the number of parking lots. Regarding the Bingle Concept, the angle, length, width, and access lane of any parking width were considered as the variables of the concept. The maximum number of parking lots was obtained by testing the different access lanes of any parking width, R and different angles, such $30°$, $45°$, and $60°$. The result showed that the most suitable selection providing the maximum number of parking spaces was the $60°$ of angle and 22.66 feet of R. In the future, researchers might study additional parking width and length to obtain more parking space, thus determining the trade-offs in space utilization, accessibility, and safety.

Keywords: Parking, Optimizing, Bingle Concept

Introduction

Parking is an activity where motor vehicles are stopped temporarily in a specific location. According to Zafirah et al. (2023), a parking lot, often known as a car park, is where individuals can park their vehicles. Parking systems and arrangements can vary from on-street parking, parking in public areas, and parking in multi-story parking buildings. A narrow and limited space is often the result of irregular urban growth, hilly terrain, and dense populations. Ogata and Oya (2023) stated that the narrow area limits the space available for the parking lot. This results in a narrow parking area for one car because of the increasing number of parking lots.

Parking can be challenging when parking is limited while the number of vehicles continues to increase. Therefore, one of the effects of this parking problem is road congestion. Qi et al. (2023) stated in their article that the average time spent by drivers in Kuala Lumpur is 25 minutes looking for a free parking lot. Vehicles frequently use the same path, searching for an open parking space when parking is scarce and full, resulting in traffic jams on the surrounding roads and the parking area itself. The increasing demand for parking spaces, particularly in urban areas, has created significant challenges in optimizing the use of narrow parking spaces.

Bingle et al. (1987) stated that access lane width enhances visibility, makes it easier for drivers to enter and exit, and reduces incidents. The parking angle selection depends on space availability, traffic conditions, and convenience. The Bingle concept might reduce the parking

constraints while allowing maximum driver convenience. Besides that, it can optimize multiple parking spaces on a specific site. The FSKM UiTM Seremban 3 met all requirements regarding this project that are to be selected as the main scope of this project. To improve space efficiency, traffic diversion, safety, and parking experience, the project incorporates different angles and widths of access lanes for parking.

Methodology

This section indicates the method implemented in this project, including data collection and model formulation. The collected data was analysed to assess the parking type, optimize the number of parking lots, and validate the mathematical model.

Data including the total area FSKM UiTM Seremban 3, A_T , length, L, width, W, access lane's width, R, angle, θ, of the parking lots, and the number of parking lots was gathered. The collected data must be appropriate for the details necessary to implement the mathematical structure.

The Mathematical Model for Parking Lot at FSKM UiTM Seremban 3

A mathematical model was applied to optimise the number of parking spaces. This model considers two distinct parking types: standard car parking, known as Type 1, and disabled parking (OKU), known as Type 2. The total area of each parking type was set as A_1 and A_2 .Meanwhile, the total area of FSKM UiTM Seremban 3 was A_T .

Number of Parking Spaces

The total area is occupied by two different parking types which are A_1 and A_2 can be used to obtain the number of parking lots. The length, width, angle and access lane of any width were the significant factors that influenced the total area for both parking types. Equation 1 is the formula to calculate the number of parking lots available.

$$N_P = \frac{A_T}{A_1 + A_2}, \tag{1}$$

The total area of parking Type 1 and Type 2 can be defined as follows:

$$A_1(\theta) = W_1\left[L_1 + R_1 \cos ec\left(\theta_1\right) + \left(2W_1 - R_1\right)\cot\left(\theta_1\right)\right], \tag{2}$$

$$A_2(\theta) = W_2\left[L_2 + R_2 \cos ec\left(\theta_2\right) + \left(2W_2 - R_2\right)\cot\left(\theta_2\right)\right], \tag{3}$$

Where,

N_P = Number of Parking Lot
A_T = Total Area of FSKM UiTM Seremban 3
A_1 = Total Area for Parking Type 1
A_2 = Total Area for Parking Type 2

L_1 = Length for Parking Type 1
L_2 = Length for Parking Type 2
W_1 = Width for Parking Type 1
W_2 = Width for Parking Type 2
θ_1 = Angle for Parking Type 1
θ_2 = Angle for Parking Type 2
R_1 = Access lane for Parking Type 1
R_2 = Access lane for Parking Type 2

Data Collection

The plan of FSKM UiTM Seremban 3 can be illustrated in Figure 1. There is only a single type of angle parking accessible at FSKM UiTM Seremban 3, which is $90°$ Parallel parking. Different parking kinds will have varied measurements even if all parking lots have the same angle.

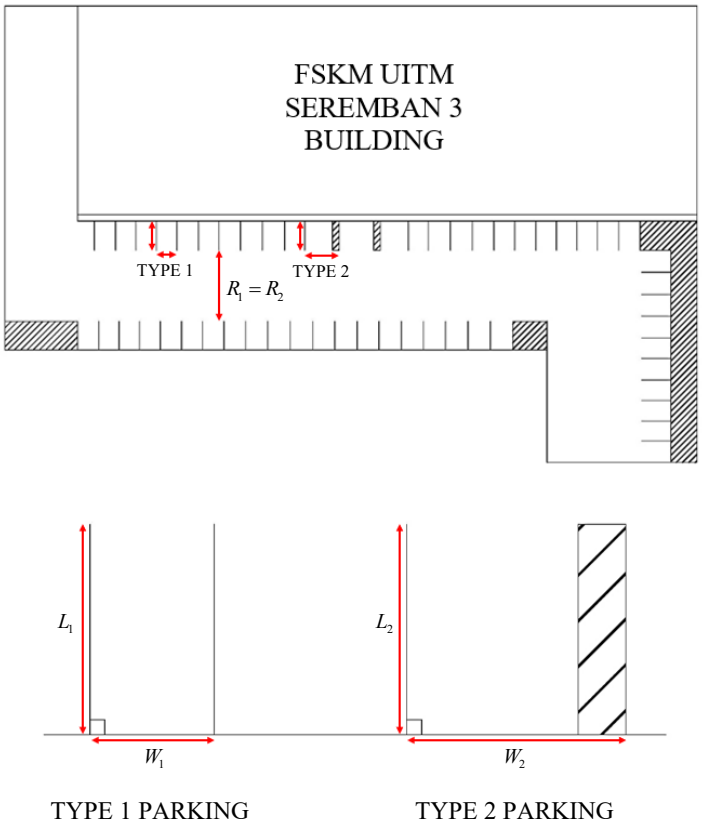


Figure 1: The Plan of FSKM UiTM Seremban 3

By referring to Figure 1, there are two types of parking lots. Type 1 is the standard type of car park, better known as the standard car parking type. For Type 1 parking, the measurements for length, width and access lane for any width are labelled as L_1, W_1 and R_1. As for Type 2 parking, this parking is disabled parking. For this type of parking, the measurements for length, width and access lane for any width are labelled as L_2 , W_2 and R_2.

From the parking layout in Figure 1, Type 1 appears to be more plentiful than Type 2 parking. This is because the demand for parking lots for regular accessible vehicles is usually greater than that for parking lots for people with disabilities. Furthermore, parking Type 2 is wider than parking Type 1 to allow wheelchair access and other mobility devices. Naturally, Type 2 occupies more space than standard parking, so the number of car parks that can be created will drop.

Data including the total area of FSKM UiTM Seremban 3, A_T , length, L, width, W, access lane's width, R, angle, θ of the parking lot, and the number of parking was gathered from the Facility Center of UiTM Seremban 3. The collected data must be appropriate for the details necessary to implement the mathematical structure. Table 1 below displays the gathered data.

Table 1: Data Collection of FSKM UiTM Seremban 3 Parking Lots

Information	**Type of Parking**	**Value (ft)**
Total area of FSKM UiTM Seremban 3, A_T		41020.02
The total area of FSKM UiTM Seremban 3's building		25510.30
Total area of parking lot space (54 spaces – standard car) (2 spaces – OKU)		15924.25
Length of parking lots, L	Standard car	15.91
	OKU	15.91
Width of parking lots, W	Standard car	7.56
	OKU	11.14
An access lane of any width, R	Standard car	23.16
	OKU	23.16

Results and Discussions

This project has identified the number of parking lots based on different angles and access lanes of any width. This is because this project wants to determine which angle and value of R can produce more parking lots than the existing car park at FSKM UiTM Seremban 3. Tables 2 and 3 show all the angles and new R-values that can be used to determine the current number of parking lots. The new R-values were reduced by 0.5 feet from the original, which is 23.16. The angles 30°, 45° and 60° of parking lots are the standard angles in most urban areas and parking guidelines provided by Dewan Bandaraya Kuala Lumpur.

Table 2: The Number of Parking Lots for Each New *R*-value by Each Angle for Case 1 to 3

Case	New *R*-values for Both Types of parking (ft)	Angle, θ for Both Types of parking	Area for Type 1 parking (ft)	Area for Type 2 parking (ft)	Total Area Occupied (ft)	Total Number of Parking
1	22.66	30°	364.17	674.77	1038.94	39
	22.66	45°	305.55	530.00	835.54	49
	22.66	60°	285.18	466.28	751.46	55
2	22.16	30°	363.15	673.28	1036.43	40
	22.16	45°	303.98	527.69	831.67	49
	22.16	60°	283.00	463.06	746.06	55
3	21.66	30°	362.14	671.79	1033.93	40
	21.66	45°	302.41	520.77	823.18	50
	21.66	60°	280.82	453.41	734.23	56

Table 3: The Number of Parking Lots for Each New *R*-value by Each Angle for Case 4 to 6

Case	New *R*-values for Both Types of parking (ft)	Angle, θ for Both Types of parking	Area for Type 1 parking (ft)	Area for Type 2 parking (ft)	Total Area Occupied (ft)	Total Number of Parking
4	21.16	30°	361.13	670.29	1031.42	40
	21.16	45°	300.85	523.08	823.92	50
	21.16	60°	278.63	456.63	735.26	56
5	20.66	30°	360.12	668.80	1028.92	40
	20.66	45°	299.28	520.77	820.05	50
	20.66	60°	276.45	453.41	729.86	56
6	20.16	30°	359.10	667.31	1026.41	40
	20.16	45°	297.72	518.46	816.18	50
	20.16	60°	274.27	450.20	724.47	57

Based on Table 6, when the R-value is reduced to 20.16 feet, and the angle changes to $60°$, it has the largest number of parking spaces, which is 57. This number of parking spaces has also exceeded the existing number of parking spaces at FSKM UiTM Seremban 3, which has 56 parking lots.

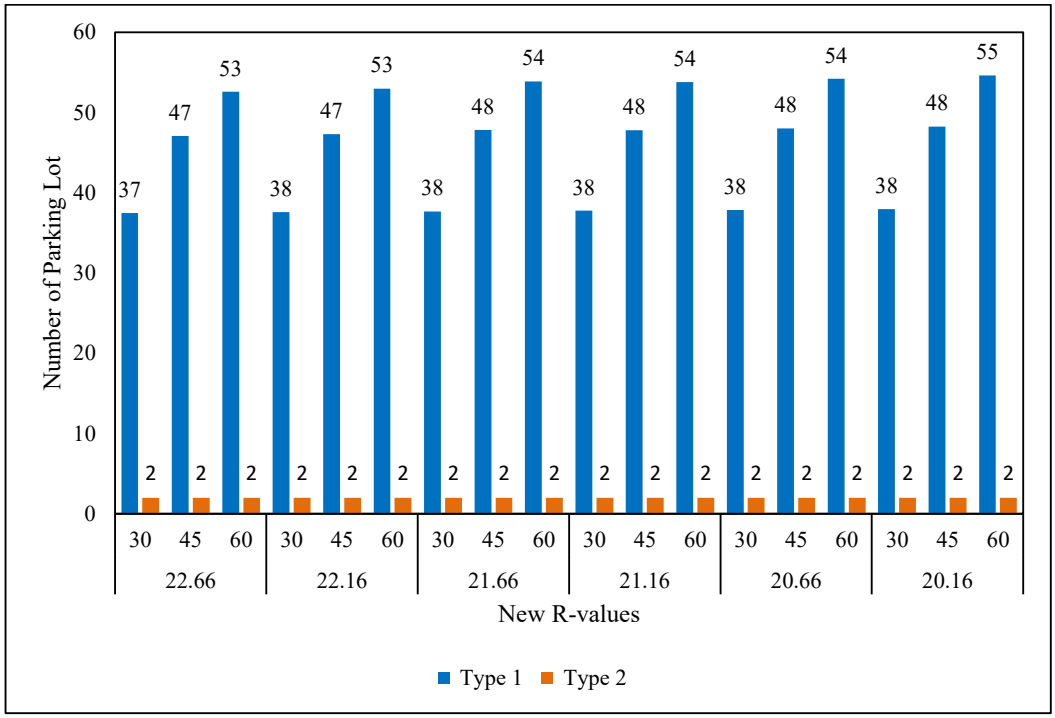


Figure 2: Number of Parking Lots for Each New *R*-value by Each Angle

The graph in Figure 2 illustrates the correlation between the number of parking lots and the new *R*-values by each angle. Six different new R-values have been calculated for each angle. It shows that the change of *R* to 20.16 and the angle to $60°$ The parking produces the optimal number of parking lots, which is 55 for Type 1 parking and 2 for Type 2 parking.

Therefore, the objective of this project, which is to optimize the number of parking spaces, has been achieved. The best angle for parking lot design is $60°$ while the best access lane of any width, R is 20.16 ft. Thus, it is proven in this project that the angle, θ and access lanes of any width, R play a vital role because changing these two factors can increase the number of parking lots in a place. The results demonstrated high accuracy, indicating that the model from the Bingle article is appropriate for this project.

Conclusion

In conclusion, by examining the relationship between variations in parking lot angle, access lane of any width and the number of parking lots, this project offers insightful information for improving parking lot design. This study aims to optimise FSKM UiTM Seremban 3's parking lot design to handle the growing number of cars and narrow space problems. To determine the optimal parking lot maximisation number, this project applies the Bingle concept and considers variables like parking angle, access lane's width, length, width and total area. Excel was used to analyse the data. The result underlines how crucial it is to balance security, visibility, and capacity. This project offers practical methods to increase parking utilisation and reduce city congestion.

Future research may apply this mathematical model based on the established parking guidelines, as Dewan Bandaraya Kuala Lumpur (DBKL) has provided specific guidelines for parking spot design, such as length, width, and access lane width. Furthermore, future research can aim to implement this mathematical model over a more extended area. The optimal optimization in a parking lot depends on choosing the right angle. It's also influencing the width of the access lane. The length and width of the parking lot can be modified to create more parking spaces that fit the area.

Acknowledgement

The authors express gratitude to all reviewers and anonymous referees for their valuable comments and recommendations, which enhanced the presentation of this article.

References

Bingle, R., Meindertsma, D., Oostendorp, W., & G. Klaasen. (1987). Designing the optimal placement of spaces in a parking lot. *Mathematical Modelling*, 9(10), 765–776. https://doi.org/10.1016/0270-0255(87)90047-9

Zafirah, N. M. S., Rahimah, S. B., & Najir, M. T. (2023). A Mathematical Model for The Optimization of Parking Lot Design. *Mathematics Letters*, 2(2), 239–247.

Ogata, R., & Oya, H. (2023). A Parking Path Planning Method Based on Circular Arcs and Lines. *Proceedings of the International Conference on New Trends in Applied Sciences*, 1, 111–116. https://doi.org/10.58190/icontas.2023.63

Qi, L. Z., Tien, D. T. K., & King Phang, S. (2023). Integrated smart public parking system for Malaysia. *Journal of Physics: Conference Series,* 2523(1). https://doi.org/10.1088/1742-6596/2523/1/012040

Numerical Solution of Unsteady Boundary Layer Flow Over a Stretching Sheet in a Porous Medium Using BVP4C

Nurul Ain Ashiqin binti Ahmad Tarmizi[1,*] and Dr Siti Hidayah binti Muhad Saleh[1]
[1] *College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 70300, Seremban, Negeri Sembilan, Malaysia*
[*]2023389991@student.uitm.edu.my

Abstract

The movement of fluid within a porous medium plays a pivotal role in various industrial applications including water circulation in geothermal reservoirs, nuclear waste storage, food processing and others. Nonetheless, examining unsteady boundary layer flow in a stretching sheet over a porous medium presents a complex interplay of nonlinear differential equations influenced by multiple external factors. This study also incorporates the effects of velocity and thermal slip to provide deeper insights. By applying similarity transformation, the governing non-linear partial differential equations were transformed into a system of nonlinear ordinary differential equations. Subsequently, the ordinary differential equations were numerically solved using the BVP4C method within MATLAB software. Numerical results are presented for the dimensionless velocity profile, temperature profile, skin friction coefficients and the local Nusselt number under varying parameter conditions. Following that, the results were validated by comparing the BVP4C method with other established numerical methods ensuring the reliability of the findings. The findings indicate that an increase in the permeability parameter leads to a decreased velocity profile and skin friction coefficient while simultaneously enhancing the temperature profile and the local Nusselt number.

Keywords*:* unsteady flow, porous medium, stretching sheet, BVP4C

1. Introduction

The theory of boundary layer flow was first introduced by Ludwig Prandtl who is German scientist [1]. The study was a significant development in understanding the fluid behaviour around solid object since he has made it necessary to analyze more simply. This flow can be categorized into steady and unsteady states depending on whether the fluid properties remain constant or change over time. Therefore, this study focuses on unsteady boundary layer flow and heat transfer in a porous medium that may be found in nature such as soil, road pavement, petroleum reservoirs, biological tissues, print paper, wood, and nanostructured materials [2].

To analyze unsteady boundary layer flow over porous media, numerical techniques such as the BVP4C method are employed. BVP4C stands for boundary value problem for ordinary differential equations with fourth-order accuracy is a reliable numerical solver that handles complex fluid flow problems with boundary conditions specified at multiple points. This method is particularly useful for higher-order differential equations that do not have closed-form analytical solutions, making it widely applicable in engineering, physics, and biological studies [3]. Researchers have extensively used BVP4C to investigate various fluid dynamics problems, including unsteady magnetohydrodynamic (MHD) flows, nanofluid interactions, and convective heat transfer in porous media as seen in [4] and [5].

Numerous studies have examined unsteady MHD flow over porous surfaces using different analytical and numerical techniques. From [6] and [7], they analyzed Casson fluid over a vertical porous plate using the Galerkin finite element method and Laplace transform method, respectively, highlighting how heat source parameters influence velocity and temperature profiles. Similarly, [8] studied unsteady MHD nanofluid flow over a rotating and stretchable disk using the shooting method and the fourth-order Runge-Kutta-Fehlberg method, demonstrating how magnetic parameters affect velocity and temperature. Other researchers have utilized analytical solutions, such as the optimal auxiliary function method [9] and the conformal map method [10] to study porous medium flow under various conditions.

For numerical approaches, methods like the finite difference method in [11] and [12], Fibonacci wavelet series collocation method [13] while Keller-box method [14] have been employed to investigate unsteady MHD flow in porous media. Studies have shown that fluid velocity and temperature distributions are significantly affected by parameters such as porosity, magnetic fields, and heat sources [15]. BVP4C has been extensively applied in research involving biological systems, such as nanofluid flow with injected nanoparticles in arteries [16] and heat transfer in hybrid nanoparticles within divergent channels [17].

Therefore, this study aims to further analyze on unsteady boundary layer flow over a porous stretching sheet conducted using the BVP4C method [18]. The findings will contribute to a deeper understanding of heat transfer and fluid dynamics in engineering applications, particularly in the fields of aerodynamics, energy systems, and biomedical engineering.

2. Methodology

The unsteady boundary layer flow occurred in the porous medium under the conditions of stretching sheet shown in Figure 1. The diagram applied Cartesian coordinate system with the origin marked as O was the stagnation point where the velocity of fluid was zero because of high pressure.

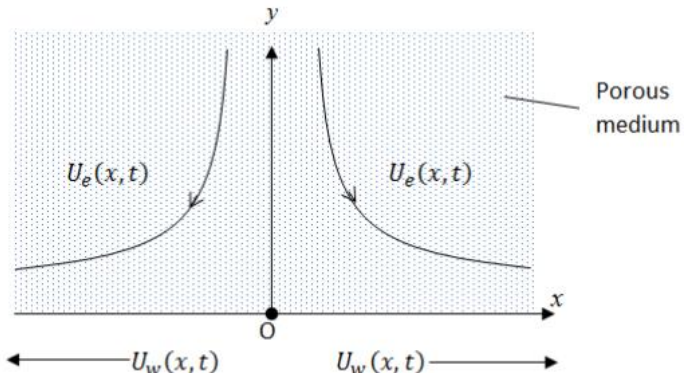


Figure 1: Physical Model of Flow and Coordinate System [18].

The arrows denoted as $U_e(x,t) = \dfrac{ax}{(1-\alpha t)}$ defined as stagnation flow velocity where $a > 0$ represent as the straining parameter and α is a constant. The direction showed that the velocity decreased as it entered the porous medium and moved away from the y-axis. Next, the arrows

represented as $U_W(x,t) = \dfrac{cx}{(1-\alpha t)}$ indicated as the stretching velocity where the $c > 0$ defined as stretching parameter. The direction displayed that the fluid flow moved along the x-axis.

The governing equations in unsteady boundary layer flow defined as the mathematical equations that represented behaviour of fluid. The equations were in the form of partial differential equations which contain more than one variable. There were three fundamental equations which are Eq. (1) - (3) and the boundary conditions in Eq. (4).

Conservation of continuity:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \tag{1}$$

Conservation of momentum:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \frac{\partial U_e}{\partial t} + U_e\frac{\partial U_e}{\partial x} + \nu\frac{\partial^2 u}{\partial y^2} + \frac{\nu}{K_1}\left(U_e - u\right), \tag{2}$$

Conservation of energy:

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \frac{k}{\rho c_p}\frac{\partial^2 T}{\partial y^2}, \tag{3}$$

Boundary conditions:

$$u = U_W(x,t) + N_1\nu\left(\frac{du}{dy}\right),$$

$$v = 0,$$

$$T = T_W + D_1\left(\frac{dT}{dy}\right) \text{ at } y = 0. \tag{4}$$

$$u \to U_e(x,t),$$

$$T \to T_\infty \text{ as } y \to \infty.$$

where:

u, v = velocity along x and y axis

T = temperature of fluid

K_1 = the permeability of the porous medium

$\nu = \frac{\mu}{\rho}$ = kinematic viscosity

μ = dynamic viscosity

ρ = density of fluid

k = thermal conductivity of fluid

t = time

c_p = specific heat.

Then, the velocity and thermal slip and were changes throughout time in the way described below in Eqs. (5) and (6)

$$N_1 = N(1-\alpha t)^{\frac{1}{2}}, \tag{5}$$

$$D_1 = D(1-\alpha t)^{\frac{1}{2}}, \tag{6}$$

where:

N = initial values of velocity

D = thermal slip factors.

Furthermore, the surface temperature was provided by Eq. (7)

$$T_W(x,t) = T_\infty + T_O\left(\frac{cx^2}{2\nu}\right)(1-\alpha t)^{\frac{1}{2}}, \tag{7}$$

where:

T_w = wall temperature

T_∞ = free stream temperature

T_o = rate of temperature increases along the sheet (constant).

Moreover, the similarity variables are provided by Eqs. (8) – (10)

$$\eta = y\sqrt{\frac{c}{v(1-\alpha t)}}, \tag{8}$$

$$\psi = y\sqrt{\frac{cv}{(1-\alpha t)}}x\, f(\eta), \tag{9}$$

$$T = T_\infty + (T_W - T_\infty)\theta(\eta), \tag{10}$$

where:

η = boundary layer thickness

ψ = stream function

T = temperature.

By substituting the similarity variables Eqs. (8) – (10) and implement the derivation process has transformed the equation Eqs. (1) – (4) into ordinary differential equations as shown in Eqs. (11) – (13)

Conservation of momentum:

$$f'''(\eta) + f(\eta)f''(\eta) - [f'(\eta)]^2 - A\left[\frac{\eta}{2}f''(\eta) + f'(\eta) - b\right] + b^2 + K\left[b - f'(\eta)\right] = 0, \tag{11}$$

Conservation of energy:

$$\frac{1}{Pr}\theta'' + f(\eta)\theta'(\eta) - 2f'\theta - \frac{A\left(3\theta + \eta\theta'\right)}{2} = 0, \tag{12}$$

where the unsteadiness parameter $A = \frac{\alpha}{c}$ and velocity ration parameter $b = \frac{\alpha}{c}$.

Boundary conditions:

$$f'(0) = 1 + \lambda f''(0),$$

$$f(0) = 0,$$

$$\theta(\eta) = 1 + \delta\theta'(\eta). \tag{13}$$

$$f'(\eta) \to b,$$

$$\theta(\eta) \to 0 \quad \text{as} \quad \eta \to \infty.$$

where $\lambda = N(cv)^{\frac{1}{2}}$ is the velocity slip parameter and $\delta = D_1\left(\frac{c}{v}\right)^{\frac{1}{2}}$ is thermal slip parameter.

After that, the physical quantity of interest that have examined in this study are the skin friction coefficient and local Nusselt number in Eqs. (14) and (15)

Skin friction coefficient:

$$C_f = \frac{\tau_w}{\rho(U_e)^2} \text{ where } \tau_w = \mu\left(\frac{\partial u}{\partial y}\right)_{y=0}, \tag{14}$$

which τ_w defined as the shear stress.

Local Nusselt number:

$$Nu_x = \frac{xq_w}{k(T_w - T_\infty)} \text{ where } q_w = -k\left(\frac{\partial T}{\partial y}\right)_{y=0}, \tag{15}$$

which q_w defined as the surface heat.

Using the similarity variables, the skin friction coefficient and local Nusselt number convert into Eqs. (16) and (17)

$$(b)^{\frac{3}{2}} C_f \left(Re_x\right)^{\frac{1}{2}} = f''(0), \tag{16}$$

$$\sqrt{b}\,Pr^{\frac{1}{2}}\,Nu_x = -\theta'(\eta)Pe_x^{\frac{1}{2}}, \tag{17}$$

where the local Reynolds number denotes as $Re_x = \frac{U_e x}{v}$, Prandtl number $Pr = \frac{\mu c_p}{k}$ and local Peclet number is $Pe_x = U_e x \frac{1}{\alpha_m}$ with α_m is the thermal diffusivity.

3. Result and Discussion

Subsequently, numerical approach using the BVP4C method is applied to solve the ordinary differential equation Eqs. (11) and (12) with the boundary conditions Eq. (13). Firstly, create a function that declare all the necessary variables that are include in the ODEs represent as the 'odefun' function and also define 'bcfun' code for the boundary conditions. Then, create an initial guess for the solution using 'bvpinit' function. This function also involves the process of discretization or separates the domain of the independent variable such as time into a mesh or grid of points. After that, call the 'bvp4c' function to solve the boundary value problem by passing the first-order ODE function, boundary conditions function and the initial guess in the form of piece-wise continuous polynomial. The following shows the transformed first order equations in Eqs. (18) – (20)

$$y_1 = f, \quad y_2 = f', \quad y_3 = f'',$$

$$y_4 = \theta, \quad y_5 = \theta'.$$

$$\frac{dy_1}{d\eta} = f' = y_2.$$

$$\frac{dy_2}{d\eta} = f'' = y_3.$$

$$\frac{dy_3}{d\eta} = f''' = -y_1 y_3 + \left[y_2\right]^2 + A\left[\frac{\eta}{2}y_3 + y_2 - b\right] - b^2 - K\left[b - y_2\right]. \tag{18}$$

$$\frac{dy_4}{d\eta} = \theta' = y_5.$$

$$\frac{dy_5}{d\eta} = \theta'' = Pr\left[-y_1 y_5 + 2y_2 + \frac{A\left(3y_4 + \eta y_5\right)}{2}\right]. \tag{19}$$

$$y_{a1} = 0, \quad y_{a2} = 1 + \lambda y_{a3}, \quad y_{a4} = 1 + \delta y_{a5}, \qquad y_{b2} = 0, \quad y_{b4} = 0. \tag{20}$$

Furthermore, the comparison of the present study was examined to ensure the validity of the numerical solution procedure used. The comparison on skin friction coefficient $f''(0)$ of steady boundary layer flow with no slip effects for different values of velocity ratio parameter b when $A = K = \lambda = \delta = 0$ are tabulated in Table 1. From this table, it indicates that a strong agreement with these methods which confirm the validation of the implemented method.

Table 1: Comparison of results of skin fiction coefficients $f''(0)$ for several values of b when $A = K = 0$ for steady flow and $\lambda = \delta = 0$ with no slip condition

b	**Nazar et al. (2004) Keller-Box Method**	**Rosali et al. (2020) Shooting Method**	**Present Result BVP4C Method**
0.01	-0.9980	-0.998065	-0.998403828
0.02	-0.9958	-0.995811	-0.996083557
0.05	-0.9876	-0.987588	-0.987729976
0.10	-0.9694	-0.969457	-0.969436241
0.20	-0.9181	-0.918118	-0.918113199
0.50	-0.6673	-0.667264	-0.667263682
2.00	2.0176	2.017503	2.017502790
3.00	4.7296	4.729282	4.729282329

Afterward, the effects of the permeability parameter K, velocity slip parameter λ and thermal slip parameter δ were investigated on velocity profile, temperature profile as well as non-dimensional quantities which are skin friction coefficients and local Nusselt number. Graphical representations of the results are used to demonstrate the behaviours of the flow and heat transfer.

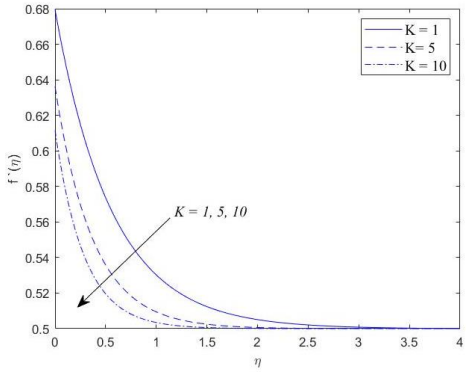


Figure 2: Velocity profile $f'(0)$ for different values of K when $Pr = 0.7$, $A = 1$, $\lambda = 1$, $\delta = 0$ and $b = 0.5$.

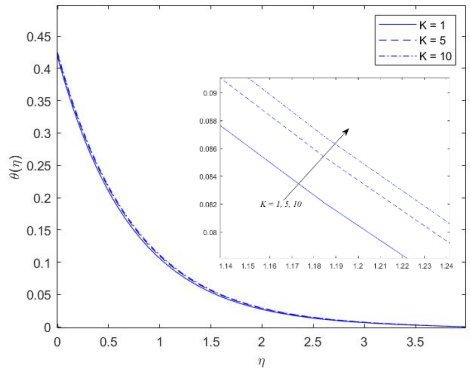


Figure 3: Temperature profile $\theta(0)$ for different values of K when $Pr = 0.7$, $A = 1$, $\lambda = 1$, $\delta = 0$ and $b = 0.5$.

Figure 2 and Figure 3 show the influence of various permeability parameter K on velocity and temperature profiles. Figure 2 indicates that the velocity decrease by the increasing values of K. Based on physical interpretation, the existence of porous medium has increases resistance to fluid flow thereby reducing the overall movement within boundary layer. From Figure 3, it is obvious show that the presence of porous medium increases the temperature profile. This indicates that the temperature decreases as the thickness of thermal boundary layer far away from the surface with the consistent heat transfer into the surrounding medium.

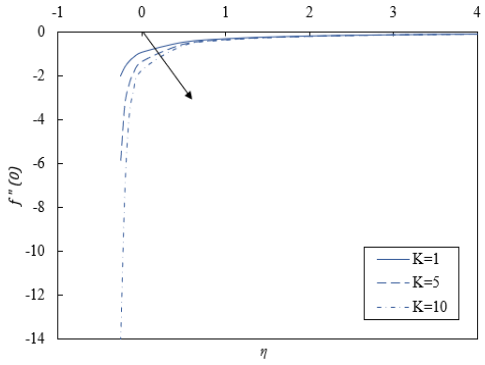


Figure 4: Skin friction coefficient $f''(0)$ for different values of K when $Pr = 0.7$, $A = 1$, $\lambda = 1$, $\delta = 1$ and $b = 0.5$.

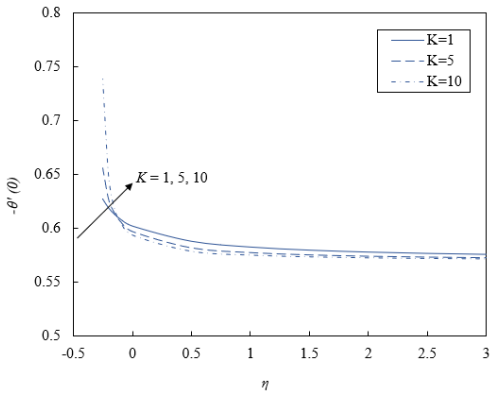


Figure 5: Local Nusselt number $-\theta'(0)$ for different values of K when $Pr = 0.7$, $A = 1$, $\lambda = 1$, $\delta = 1$ and $b = 0.5$.

Figure 4 show that the skin friction coefficients decrease as permeability parameter K increases. The values of the skin friction coefficient approach certain values with the increase in the value of velocity slip parameter λ for different values of permeability parameter K. It can be observed that the effect of permeability parameter K is small for large values of velocity slip parameter λ. Next, Figure 5 demonstrates a trend as the permeability parameter K increases, the local Nusselt number also increases. Based on physical perspective, this behaviour can be

explained by the fact that an increase in permeability enhances the fluid's ability to pass through the porous medium which increases heat transfer properties.

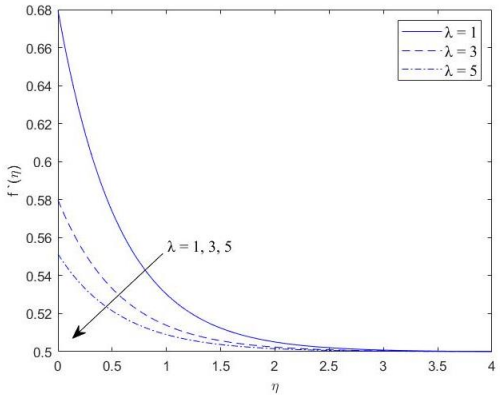


Figure 6: Velocity profile $f'(0)$ for different values of λ when Pr = 0.7, A = 1, δ = 1, b = 0.5 and K = 1.

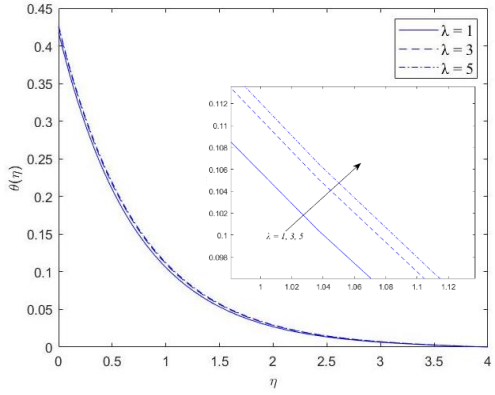


Figure 7: Temperature profile $\theta(0)$ for different values of λ when Pr = 0.7, A = 1, δ = 1, b = 0.5 and K = 1.

Figure 6 and Figure 7 show the influence of various velocity slip parameter λ on velocity and temperature profiles. Figure 6 illustrated that the fluid velocity decrease when the velocity slip parameter λ increase. This is because when velocity slip occurs, the velocity of fluid flow close to the surface is no longer equal to the velocity stretching surface. Therefore, physically this slips condition represents as a surface with partial resistance towards the fluid movement in porous medium. Moreover, the result from Figure 7 discovered that the temperature profile exhibits an increasing pattern as the values of λ increase. Higher values of λ indicate greater slip at the surface, leading to enhanced heat transfer from the surface to the fluid due to lower tangential velocity. The temperature distribution is indirectly influenced by the velocity slip parameter λ, which modifies the fluid flow close to the surface.

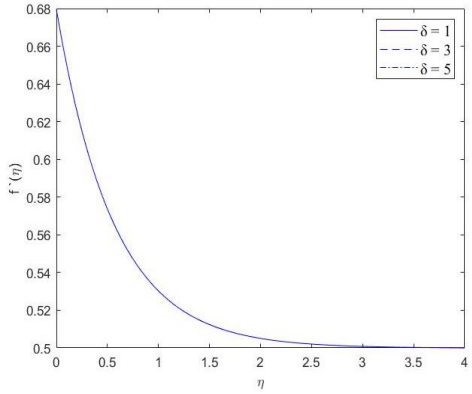


Figure 8: Velocity profile $f'(0)$ for different values of δ when Pr = 0.7, A = 1, b = 0.5, K = 1 and λ = 1.

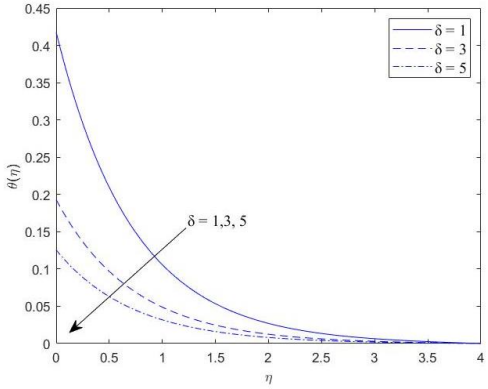


Figure 9: Temperature profile $\theta(0)$ for different values of δ when Pr = 0.7, A = 1, b = 0.5, K = 1 and λ = 1.

Figure 8 and Figure 9 show the influence of various thermal slip parameter δ on velocity and temperature profiles. From Figure 8, it can be seen that by increasing values of the thermal slip parameter δ, the velocity profile does not change and remain constant. This shows that the thermal slip parameter has no observable effect on the velocity profile. This could occur when momentum diffusion dominates over thermal effects, meaning the velocity field is primarily governed by the momentum boundary conditions rather than thermal slip effects. Figure 9 indicates that the temperature profile decreases due to increasing values of thermal slip parameter δ. Higher δ reduces the effectiveness of heat transfer between the boundary and the fluid, leading to a thinner thermal boundary layer. Physically, this means that the temperature changes more rapidly near the boundary and the thermal influence does not extend far into the fluid.

4. Conclusion

This study investigates the unsteady boundary layer flow of a stretching sheet over a porous medium. The boundary layer flow equations in nonlinear partial differential equations have been transformed into dimensionless ordinary differential equations through similarity transformation. Then, the equations are solved numerically using BVP4C method in MATLAB software. Hence, the results that are obtained as follows:

i. The velocity profile $f'(0)$ decrease as the increasing values of permeability parameter K and velocity slip parameter λ.

ii. The velocity profile $f'(0)$ constant when thermal slip parameter δ increases.

iii. The temperature profile $\theta(0)$ decrease with the increasing values of permeability parameter K and thermal slip parameter δ.

iv. The temperature profile $\theta(0)$ increase when velocity slip parameter λ increases.

v. The skin friction coefficient $f''(0)$ decreases due to increasing permeability parameter K.

vi. The skin friction coefficient $f''(0)$ increases with the increasing velocity slip parameter λ.

vii. The local Nusselt number $-\theta'(0)$ increase when permeability parameter K increase.

viii. The local Nusselt number $-\theta'(0)$ decrease as velocity slip λ and thermal slip parameter δ increases.

Acknowledgments

The authors would like to express their sincere gratitude to Universiti Teknologi MARA (UiTM) for providing the necessary facilities and resources to conduct this research. Special appreciation is extended to the College of Computing, Informatics and Mathematics for their continuous support and valuable guidance.

References

[1] Tani, I. (1977). History of boundary layer theory. *Annual Review of Fluid Mechanics*, *9*(1), 87–111. https://doi.org/10.1146/annurev.fl.09.010177.000511

[2] Kamrava, S., Sahimi, M., & Tahmasebi, P. (2021). Simulating fluid flow in complex porous materials by integrating the governing equations with deep-layered machines. Npj Computational Materials, 7(1). https://doi.org/10.1038/s41524-021-00598-2

[3] Sudarmozhi, K., Iranian, D., & Alessa, N. (2024). Investigation of melting heat effect on fluid flow with brownian motion/thermophoresis effects in the occurrence of energy on a stretching sheet. *Alexandria Engineering Journal /Alexandria Engineering Journal*, *94*, 366–376. https://doi.org/10.1016/j.aej.2024.03.065

[4] Zhou, J., Abidi, A., Shi, Q., Khan, M. R., Rehman, A., Issakhov, A., & Galal, A. M. (2021). Unsteady radiative slip flow of MHD Casson fluid over a permeable stretched surface subject to a non-uniform heat source. *Case Studies in Thermal Engineering*, *26*, 101141. https://doi.org/10.1016/j.csite.2021.101141

[5] Verma, A. K., Rajput, S., Bhattacharyya, K., & Chamkha, A. J. (2022). Nanoparticle's radius effect on unsteady mixed convective copper-water nanofluid flow over an expanding sheet in porous medium with boundary slip. *Chemical Engineering Journal Advances*, *12*, 100366. https://doi.org/10.1016/j.ceja.2022.100366

[6] Goud, B. S., Kumar, P. P., & Malga, B. S. (2020). Effect of Heat source on an unsteady MHD free convection flow of Casson fluid past a vertical oscillating plate in porous medium using finite element analysis. *Partial Differential Equations in Applied Mathematics*, *2*, 100015. https://doi.org/10.1016/j.padiff.2020.100015

[7] Rao, S. R., Vidyasagar, G., & Deekshitulu, G. (2021). Unsteady MHD free convection Casson fluid flow past an exponentially accelerated infinite vertical porous plate through porous medium in the presence of radiation absorption with heat generation/absorption. *Materials Today: Proceedings*, *42*, 1608–1616. https://doi.org/10.1016/j.matpr.2020.07.554

[8] Mandal, S., & Shit, G. (2021). Entropy analysis on unsteady MHD biviscosity nanofluid flow with convective heat transfer in a permeable radiative stretchable rotating disk. *Chinese*

Journal of Physics/Zhōngguó Wùlǐ Xuékān, *74*, 239–255. https://doi.org/10.1016/j.cjph.2021.07.036

[9] Marinca, B., Herişanu, N., & Marinca, V. (2023). Analytical solutions for solving unsteady flow of gas through a porous medium by using auxiliary functions method. *Journal of Computational and Applied Mathematics*, *432*, 115296. https://doi.org/10.1016/j.cam.2023.115296

[10] Almalki, F. M. K., & Meylan, M. H. (2024). Analytic solution of the free boundary problem for porous media flow using a conformal map validated by the boundary element method. *Engineering Analysis with Boundary Elements*, *166*, 105806. https://doi.org/10.1016/j.enganabound.2024.105806

[11] Fahim, M., Sajid, M., Ali, N., & Naveed, M. (2023). Unsteady blood flow of Carreau fluid in a porous saturated medium with stenosis under the influence of acceleration and magnetic fields: A comprehensive analysis. *Computers in Biology and Medicine*, *164*, 107278. https://doi.org/10.1016/j.compbiomed.2023.107278

[12] Goud, B. S., Srilatha, P., Srinivasulu, T., Reddy, Y. D., & Kumar, K. S. (2023). Induced by heat source on unsteady MHD free convective flow of Casson fluid past a vertically oscillating plate through porous medium utilizing finite difference method. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2023.01.378

[13] Preetham, M., Kumbinarasaiah, S., & Alshehri, M. (2024). Application of Fibonacci wavelet in the analysis of unsteady MHD Williamson nanofluid flow over a permeable stretching sheet via porous medium. *Results in Physics*, 107853. https://doi.org/10.1016/j.rinp.2024.107853

[14] Khan, D., Ali, G., Kumam, P., Sitthithakerngkiet, K., & Jarad, F. (2023). Heat transfer analysis of unsteady MHD slip flow of ternary hybrid Casson fluid through nonlinear stretching disk embedded in a porous medium. *Ain Shams Engineering Journal/Ain Shams Engineering Journal*, *15*(2), 102419. https://doi.org/10.1016/j.asej.2023.102419

[15] Zegeye, G. B., Haile, E., & Awgichew, G. (2024). Viscous dissipation and Joule heating effects of Carreau nanofluid axisymmetric flow past unsteady radially stretching porous disk. *International Journal of Thermofluids*, *22*, 100655. https://doi.org/10.1016/j.ijft.2024.100655

[16] Waqas, H., Farooq, U., Liu, D., Alghamdi, M., Noreen, S., & Muhammad, T. (2022). Numerical investigation of nanofluid flow with gold and silver nanoparticles injected inside a stenotic artery. *Materials & Design*, *223*, 111130. https://doi.org/10.1016/j.matdes.2022.111130

[17] Zada, L., Ullah, I., Alqahtani, A. M., Nawaz, R., Khan, H., & Alam, K. (2024). Enhancing energy efficiency and heat transfer performance of engine oil flow through hybrid nanoparticles in convergent/divergent channel. *Results in Engineering*, *22*, 102027. https://doi.org/10.1016/j.rineng.2024.102027

[18] Rosali, H., Badlilshah, M. N., Johari, M. a. M., & Bachok, N. (2020). Unsteady Boundary Layer Stagnation Point Flow and Heat Transfer over a Stretching Sheet in a Porous Medium with Slip Effects. *CFD Letters*, *12*(10), 52–61. https://doi.org/10.37934/cfdl.12.10.5261

SELECTION OF SMARTPHONE BRANDS BY USING FUZZY TOPSIS APPROACH

Nurin Hannani Binti Halimi[1]*, Nor Hilaliyah Binti Mohd Jamil[2] and Siti Nazifah Binti Zainol Abidin[3]

[1,2,3]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia.

*hannaninurin5@gmail.com

Abstract

This study investigates the selection of smartphone brands using the Fuzzy TOPSIS approach, a multi-criteria decision-making method. A smartphone is an essential need, thus choosing the best option among various brands is a crucial process for consumers. Five criteria; price, RAM, storage, battery and camera were used to evaluate five popular smartphone brands; Samsung, Apple, Oppo, Vivo and Huawei. A questionnaire was assessed by a committee of ten decision makers from diverse backgrounds and the linguistic variables used were converted to triangular fuzzy numbers. The Fuzzy TOPSIS steps involve constructing decision matrices, determining the weight of criteria, normalizing values and calculating distances from ideal solutions. The findings ranked the brands, with Apple as the most preferred, which has a closeness coefficient of 0.689. This study demonstrates the effectiveness of Fuzzy TOPSIS in handling ambiguous data, optimizing decisions, and providing insights for consumers and companies to enhance product selection and marketing strategies in complex environments.

Keywords: Smartphone, consumers, MCDM, Fuzzy TOPSIS, decision makers

1. Introduction

The sophistication of technology today is aimed at easing people's daily lives, including the development of communication technologies. The first telephone was invented in the 1980s and continues to evolve to this day [7]. Now, smartphones are one of the most essential electronic devices in daily life. Smartphones can be defined as communication tools with computer-like features and functionality [14]. Smartphones have become a worldwide phenomenon. In today's world, the most well-known smartphone brands are Apple, Samsung, Oppo, Vivo, Realme and others [14].

Multi-Criteria Decision Making (MCDM) is a method of assessing several criteria and selecting the best alternative among the alternatives provided [13]. MCDM can be applied to daily life problems as it involves the decision-making process, which is a common problem in human life. Today, complex decision-making problems can be addressed by using mathematical equations, economic theories, advanced statistics and computer technology that will calculate and estimate solutions automatically [15].

Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is the classical method of MCDM where decision alternatives are ranked and the alternative which is the closest to the positive ideal solution (PIS) and the farthest to the negative ideal solution (NIS) are chosen [10]. Fuzzy TOPSIS is an extended version of the classical TOPSIS approach. Fuzzy set theory was applied to problems that are unclear and ambiguous [1]. When there was a lack of accurate data or when there were subjective judgments involved, it was very helpful to use the Fuzzy TOPSIS approach.

Nowadays, every individual needs to have a smartphone as one of the essential electronic devices in daily life. However, the variety of smartphones in the marketplace makes it difficult for consumers to choose the most suitable smartphone for their daily use. In addition, consumers also undergo difficulty

in determining the criteria in choosing a smartphone because there are too many criteria to consider before buying a smartphone, such as price, random access memory (RAM), processor and other criteria [6]. On the other hand, there are many Multi-Criteria Decision Making (MCDM) approaches that can be applied to the selection of smartphone brands.

There are three objectives involved in this study. Firstly, this study needs to identify the main criteria that influence the selection of smartphone brands. Next, this study needs to apply the Fuzzy TOPSIS approach for the selection of smartphone brands using the main criteria found in the first objective. Lastly, this study aims to analyze the most preferred smartphone brands using the Fuzzy TOPSIS approach.

The study used the most suitable MCDM approach that will provide a better understanding in the selection of smartphone brands under a fuzzy environment. Firstly, this study allows the decision-making process in a complex market. Consumers can make the best choice of a smartphone when they focus on several main criteria and alternatives. Next, this study helps in handling imprecise and uncertain information. Ambiguous data can be calculated by applying the Fuzzy TOPSIS approach and obtaining the desired results. Furthermore, this study will optimize multi-criteria decision-making processes. Consumers will be able to minimize the cost of purchasing, but at the same time obtain the best choices among others by focusing on the main criteria.

This study is to observe the preferences of consumers among smartphone brands in the marketplaces. The committee of ten decision makers are formed, which consists of smartphone promoters, smartphone repair technicians, smartphone sellers and experienced consumers of different smartphone brands. The main criteria chosen were price, RAM, storage, battery and camera. The smartphone brands used as alternatives for this study were Samsung, Apple, Oppo, Vivo, and Huawei. The most suitable MCDM approach used in this study was the Fuzzy TOPSIS approach. Therefore, the best smartphone brand among the five chosen alternatives needs to be analyzed by observing the most considered criteria using one of the approaches applied in an uncertainty environment, which is the Fuzzy TOPSIS approach.

2. Literature Review

A smartphone is a mobile phone with computer features, has an LCD touchscreen and includes software functionality such as browser and email [1]. A study by [7] has stated the idea of the first transportable telephone was the Motorola DynaTAC 8000X, which was developed in 1973 by Martin Cooper and was released ten years later in 1983. According to [7], as technology advanced, mobile phones evolved into smartphones, with Apple creating the first touchscreen smartphone in 2007. Now, smartphones in the marketplaces are using touchscreens that are equipped with software and apps and are much easier to use.

There are many criteria that consumers should consider when selecting a smartphone. There are five common criteria in selecting smartphone brands, which are price, RAM, storage, battery and camera. There are five main criteria used in [5] to determine the best smartphone models among ten smartphone models by using Fuzzy-AHP and PROMETHEE methods. In addition, a study by [2] has used all five main criteria in their study to evaluate the best smartphone brands among consumers.

The Fuzzy TOPSIS approach is an extension of TOPSIS, which involves solving decision-making problems by applying the concept of FST to the problem. Based on a study by [4] on selecting the most suitable candidate to fulfil the system analysis engineer position, linguistic variables were applied to the solutions and then the FPIS and FNIS were determined. Many studies were conducted using Fuzzy TOPSIS, especially when the problems involved an uncertain environment and imprecise data.

In the context of selecting smartphone brands, there are several studies that apply the Fuzzy TOPSIS approach. The Fuzzy TOPSIS approach has been used in a study by [11] regarding smartphone selection and the study resulted in the Q-Mobile smartphone brand being the best smartphone among six alternatives. The criteria used in this study were RAM, ROM, processor, camera, display size, model and price. In addition, [3] has applied the Intuitionistic Fuzzy Set (IFS) to the TOPSIS approach in the selection of a smartphone. The criteria considered in this study were battery backup-related criteria, weight-related criteria, camera-related criteria, cost/price-related criteria and looks and design-related criteria. The study resulted in alternative 3 being the most selected smartphone among the four smartphones.

3. Methodology

3.1. Mathematical Concept and Definitions

3.1.1. Fuzzy Set Theory (FST)

The definition of FST is referred as a set of fuzzy numbers in the universal set of U [4]. The idea of FST was introduced by Zadeh in 1965, where it involves solving the situations that are highly complex using fuzzy numbers instead of crisp values. Let U be the universal set, a fuzzy set A in U can be described with a membership function of $\mu_A(x)$ which takes the values in interval [0,1]. The closer the value to 1, the higher the degree of membership of the element x in the fuzzy set A. Other than that, a fuzzy set A in universal set U is said to be normal if the fuzzy set A satisfy $\exists x_i \in U, \mu_A(x_i) = 1$. Based on Figure 1, it is explained that a fuzzy subset in universal set U is defined as fuzzy number A where the membership function is convex and normal [4].

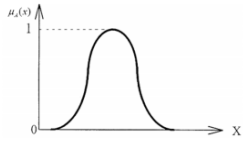


Figure 1: A fuzzy number A

3.1.2. Triangular Fuzzy Number (TFN)

Triangular Fuzzy Number (TFN) is one of the membership functions that can be used to represent linguistic variables [12]. There are several types of membership functions, which include triangular, trapezoidal, Gaussian and quadrilateral [12]. Figure 2 shows the triangular fuzzy number A, which consists of (a_1, a_2, a_3).

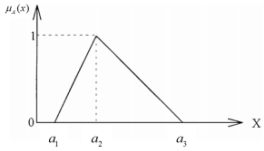


Figure 2: A triangular fuzzy number A

3.1.3. Operations of Triangular Fuzzy Number

Let A and B be two fuzzy numbers expressed in intervals $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$ and these are defined as real numbers. There are four main operations which involve two positive fuzzy numbers represented as A and B [8].

a) Addition:

$$A + B = (a_1 + b_1, a_2 + b_2, a_3 + b_3) \tag{1}$$

b) Multiplication:

$$A \cdot B = (a_1 \cdot b_1, a_2 \cdot b_2, a_3 \cdot b_3) \tag{2}$$

c) Subtraction:

$$A - B = (a_1 - b_1, a_2 - b_2, a_3 - b_3) \tag{3}$$

d) Division:

$$\frac{A}{B} = \left(\frac{a_1}{b_3}, \frac{a_2}{b_2}, \frac{a_3}{b_1}\right) \tag{4}$$

3.2. Implementation of Fuzzy TOPSIS Approach

There are 9 main steps in using this approach, which was obtained from a previous study by [4]. The significant results are explained in the steps of the Fuzzy TOPSIS approach.

Step 1: The decision makers are identified and the criteria are determined.
A committee of ten decision makers has been formed. The decision makers are assigned from D1 to D10. Several alternatives and criteria to evaluate smartphone brands are determined. The alternatives and criteria chosen are as in Table 1.

Table 1: List of Criteria and Alternatives

Criteria		**Alternatives**	
C1	Price	A1	Samsung
C2	RAM	A2	Apple
C3	Storage	A3	Oppo
C4	Battery	A4	Vivo
C5	Camera	A5	Huawei

Step 2: The suitable linguistic variable for the weight of criteria and rating of alternatives are determined.
Linguistic variables were used to represent the scale of ratings. The linguistic variables then were converted into TFN since this study used triangular membership functions. The linguistic variables used for the weight of criteria and rating of alternatives are as in Table 2 and Table 3. The linguistic variables are derived from [4].

Table 2: Linguistic variables for the weight of criteria [4]

Linguistic Variables	**Fuzzy Numbers**
Very Low (VL)	(0, 0, 0.1)
Low (L)	(0, 0.1, 0.3)
Medium Low (ML)	(0.1, 0.3, 0.5)
Medium (M)	(0.3, 0.5, 0.7)
Medium High (MH)	(0.5, 0.7, 0.9)
High (H)	(0.7, 0.9, 1.0)
Very High (VH)	(0.9, 1.0, 1.0)

Table 3: Linguistic variables for rating of alternatives [4]

Linguistic Variables	**Fuzzy Numbers**
Very Poor (VP)	(0, 0, 1)
Poor (P)	(0, 1, 3)
Medium Poor (MP)	(1, 3, 5)
Fair (F)	(3, 5, 7)
Medium Good (MG)	(5, 7, 9)
Good (G)	(7, 9, 10)
Very Good (VG)	(9, 10, 10)

Step 3: The weight of criteria and rating of alternatives are aggregated and the aggregated fuzzy weight of criteria and rating of alternatives are obtained.
The weight of criteria and rating of alternatives are aggregated to obtain the aggregated fuzzy weight, w_j of criteria, C_j and aggregated fuzzy rating, x_{ij} of alternatives, A_i. Let N be the group of decision

makers, the aggregated fuzzy weight of criteria and aggregated fuzzy rating of alternatives can be calculated using (5) and (6) respectively.

$$\text{Weight of criteria, } w_j = \frac{1}{N}\left[w_j^1 + w_j^2 + \cdots + w_j^N \right] \tag{5}$$

$$\text{Rating of alternative, } x_{ij} = \frac{1}{N}\left[x_{ij}^1 + x_{ij}^2 + \cdots + x_{ij}^N \right] \tag{6}$$

where w_j^N and x_{ij}^N are the weight of criteria and rating of alternatives respectively of the Nth decision makers. The weight of criteria and rating of alternatives of the decision makers then can be expressed in matrix form as (7) and (8) respectively.

$$W = \left[w_1, w_2, \ldots, w_n \right] \tag{7}$$

$$D = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{8}$$

where $x_{ij}, \forall i, j$ and $w_j, j = 1, 2, \ldots, n$ are linguistic variables. The linguistic variables are written in the form of triangular fuzzy numbers, $x_{ij} = \left(a_{ij}, b_{ij}, c_{ij} \right)$ and $w_j = \left(w_{j1}, w_{j2}, w_{j3} \right)$. Table 4 and Table 5 show the evaluation from ten decision makers on the weight of criteria and rating of alternatives respectively. The aggregated fuzzy weight of criteria and aggregated fuzzy rating of alternatives are displayed in Table 6 and Table 7 respectively.

Step 4: The normalized fuzzy decision matrix is formed.
The normalized fuzzy decision matrix can be written as matrix R. The matrix is calculated using (9) and (10).

$$R = \left[r_{ij} \right]_{m \times n}$$

$$r_{ij} = \left(\frac{a_{ij}}{c_j^+}, \frac{b_{ij}}{c_j^+}, \frac{c_{ij}}{c_j^+} \right), j \in B; \tag{9}$$

$$r_{ij} = \left(\frac{a_j^-}{c_{ij}}, \frac{a_j^-}{b_{ij}}, \frac{a_j^-}{a_{ij}} \right), j \in C; \tag{10}$$

where c_j^+ is the maximum value of TFN and preferable for benefit criteria and a_j^- is the minimum value of TFN and preferable for cost criteria. The benefit criteria used in this study are C2, C3, C4 and C5. C1 is the cost criteria used in this study. Table 7, which is also the fuzzy decision matrix is used to calculate the normalized fuzzy decision matrix. The normalized fuzzy decision matrix is presented in Table 8.

Step 5: The weighted normalized fuzzy decision matrix is formed.
The weighted normalized fuzzy decision matrix is calculated by using (11).

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{bmatrix}, v_{ij} = r_{ij} \cdot w_j \tag{11}$$

The matrix can be formed by multiplying the normalized fuzzy decision matrix in Table 8 and aggregated fuzzy weight of criteria in Table 6. The weighted normalized fuzzy decision matrix is presented in Table 9.

Step 6: The Fuzzy Positive Ideal Solution (FPIS) and Fuzzy Negative Ideal Solution (FNIS) are computed.

The elements of the weighted normalized fuzzy decision matrix are in the interval of $[0,1]$. Therefore, the TFN of FPIS, A^+ and FNIS, A^- used in this study are as below:

$$A^+ = \left[(1,1,1),(1,1,1),(1,1,1),(1,1,1),(1,1,1)\right]$$

$$A^- = \left[(0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0)\right]$$

Step 7: The distance of FPIS and FNIS to each alternative are computed.

The distance of FPIS and FNIS are calculated using (12). Then, the summation of all distances is calculated. This step is known as the defuzzification process from three values of TFN to one crisp value. The TFN values must be divided by three to obtain the crisp value. Therefore, the total distances of FPIS and FNIS are shown in Table 10.

$$d(A,B) = \sqrt{\frac{1}{3}\left[(a_1-b_1)^2+(a_2-b_2)^2+(a_3-b_3)^2\right]} \tag{12}$$

Table 10: Total distances of FPIS and FNIS to each alternative

Alternatives	A^+	A^-
A1	1.791	3.503
A2	1.648	3.657
A3	1.869	3.406
A4	1.916	3.389
A5	1.766	3.546

Step 8: The closeness coefficient for each alternative is computed.

Table 11 indicates the closeness coefficient for each alternative. The closeness coefficient value is calculated using (13).

$$CC_i = \frac{d_i^-}{d_i^+ + d_i^-}, i = 1,2,\ldots,m \tag{13}$$

Table 11: The closeness coefficient for each alternative

CC_1	0.662
CC_2	0.689
CC_3	0.646
CC_4	0.639
CC_5	0.668

Step 9: The alternatives are ranked based on the closeness coefficient.

The closeness coefficient values are used to rank the alternatives. As the value approaches 1, it is the most preferred alternative, while if the value approaches 0, it is the least preferred alternative. In this study, the values are arranged in descending order. The highest value is considered the most preferred alternative, while the lowest value is considered the least preferred alternative. Therefore, the ranking of alternatives is A2, A5, A1, A3 and A4.

Table 4: Evaluation of ten decision makers for weight of criteria

Criteria	Decision Makers									
	D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
C1	MH	MH	M	M	M	H	H	MH	VH	H
C2	MH	VH	M	H	VH	VH	H	H	VH	H
C3	VH	VH	M	VH	M	VH	H	VH	VH	H
C4	VH	VH	M	H	VH	MH	VH	VH	VH	VH
C5	MH	VH	M	VH	VH	MH	ML	MH	VH	VH

Table 5: Evaluation of ten decision makers for rating of alternatives

Criteria	Alternatives	Decision Makers									
		D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
C1	**A1**	MP	MP	P	MG	G	F	VG	G	VG	VG
	A2	P	MP	VG	F	G	F	MP	G	MG	VG
	A3	F	F	VG	G	MG	G	VG	VG	VG	MG
	A4	F	F	MP	G	MG	G	VG	VG	VG	MG
	A5	F	MP	P	MG	G	MG	G	G	G	MP
C2	**A1**	MG	VG	MP	G	MG	G	VG	G	VG	VG
	A2	VG	G	VG	VG	VG	G	VG	VG	VG	VG
	A3	MG	VG	VG	MG	MG	G	F	MG	VG	G
	A4	G	VG	F	G	MG	MG	F	G	VG	MG
	A5	G	VG	VP	VG	VG	G	VG	VG	VG	MP
C3	**A1**	VG	VG	F	G	MG	G	VG	G	VG	VG
	A2	VG	VG	F	VG	VG	G	VG	VG	VG	VG
	A3	VG	VG	F	VG	MG	G	G	VG	VG	VG
	A4	VG	VG	F	G	MG	G	G	MG	VG	VG
	A5	VG	VG	F	VG	MG	G	VG	G	VG	MP
C4	**A1**	G	VG	MG	F	VG	MG	VG	F	VG	VG
	A2	VG	MG	MG	F	MG	P	G	MG	G	VG
	A3	G	G	VG	VG	VG	G	G	VG	VG	VG
	A4	G	G	VG	MG	VG	MG	G	G	VG	VG
	A5	G	G	VG	VG	VG	G	G	VG	VG	F
C5	**A1**	VG	VG	VG	G	G	G	MG	G	VG	VG
	A2	VG	VG	VG	VG	VG	VG	VG	VG	VG	VG
	A3	VG	G	VG	MG	F	VG	MG	F	G	G
	A4	VG	VG	VG	G	F	G	MG	MG	G	G
	A5	VG	VG	VG	VG	G	VG	VG	G	G	F

Table 6: Aggregated fuzzy weight of criteria

Criteria	Weight
C1	(0.540, 0.730, 0.880)
C2	(0.720, 0.880, 0.960)
C3	(0.740, 0.880, 0.940)
C4	(0.780, 0.910, 0.960)
C5	(0.640, 0.790, 0.890)

Table 7: Aggregated fuzzy rating of alternatives

Alternatives	Criteria				
	C1	C2	C3	C4	C5
A1	(5.100, 6.700, 7.900)	(6.800, 8.400, 9.300)	(7.400, 8.900, 9.600)	(6.800, 8.300, 9.200)	(7.800, 9.300, 9.900)
A2	(4.500, 6.200, 7.600)	(8.600, 9.800, 10.000)	(8.200, 9.400, 9.700)	(5.500, 7.200, 8.600)	(9.000, 10.000, 10.000)
A3	(6.600, 8.200, 9.200)	(6.400, 8.100, 9.300)	(7.600, 9.000, 9.600)	(8.200, 9.600, 10.000)	(6.400, 8.100, 9.200)
A4	(5.800, 7.500, 8.700)	(6.000, 7.800, 9.100)	(7.000, 8.600, 9.500)	(7.400, 9.000, 9.800)	(6.800, 8.500, 9.500)
A5	(4.300, 6.200, 7.800)	(6.900, 8.100, 8.600)	(6.800, 8.300, 9.100)	(7.600, 9.100, 9.700)	(7.800, 9.200, 9.700)

Table 8: The normalized fuzzy decision matrix

Alternatives	Criteria				
	C1	C2	C3	C4	C5
A1	(0.544, 0.642, 0.843)	(0.680, 0.840, 0.930)	(0.763, 0.918, 0.990)	(0.680, 0.830, 0.920)	(0.780, 0.930, 0.990)
A2	(0.566, 0.694, 0.956)	(0.860, 0.980, 1.000)	(0.845, 0.969, 1.000)	(0.550, 0.720, 0.860)	(0.900, 1.000, 1.000)
A3	(0.467, 0.524, 0.652)	(0.640, 0.810, 0.930)	(0.784, 0.928, 0.990)	(0.820, 0.960, 1.000)	(0.640, 0.810, 0.920)
A4	(0.494, 0.573, 0.741)	(0.600, 0.780, 0.910)	(0.722, 0.887, 0.979)	(0.740, 0.900, 0.980)	(0.680, 0.850, 0.950)
A5	(0.551, 0.694, 1.000)	(0.690, 0.810, 0.860)	(0.701, 0.856, 0.938)	(0.760, 0.910, 0.970)	(0.780, 0.920, 0.970)

Table 9: The weighted normalized fuzzy decision matrix

Alternatives	Criteria				
	C1	C2	C3	C4	C5
A1	(0.294, 0.469, 0.742)	(0.490, 0.739, 0.893)	(0.565, 0.808, 0.931)	(0.530, 0.755, 0.883)	(0.499, 0.735, 0.881)
A2	(0.306, 0.507, 0.841)	(0.619, 0.862, 0.960)	(0.625, 0.853, 0.940)	(0.429, 0.655, 0.826)	(0.576, 0.790, 0.890)
A3	(0.252, 0.383, 0.574)	(0.461, 0.713, 0.893)	(0.580, 0.817, 0.931)	(0.640, 0.874, 0.960)	(0.410, 0.640, 0.819)
A4	(0.267, 0.418, 0.652)	(0.432, 0.686, 0.874)	(0.534, 0.781, 0.920)	(0.577, 0.819, 0.941)	(0.435, 0.672, 0.846)
A5	(0.298, 0.507, 0.880)	(0.497, 0.713, 0.826)	(0.519, 0.753, 0.882)	(0.593, 0.828, 0.931)	(0.499, 0.727, 0.863)

4. Results and Discussions

The results of this study provide an extensive analysis of the selection of smartphone brands using the Fuzzy TOPSIS approach. In this study, each criterion and alternative were evaluated using linguistic variables. The linguistic variables were then converted into TFN and aggregated to obtain the aggregated fuzzy weight of criteria and rating of alternatives. Then, the fuzzy decision matrix obtained is normalized. The weighted normalized fuzzy decision matrix is formed and the closeness coefficient for each alternative was calculated. The alternatives were ranked based on the closeness coefficient value. The alternative that has the closeness coefficient value approaching 1 is considered the most preferred alternative, while the value that approaches 0 is the least preferred alternative.

The ranking of alternatives is presented in Table 12 and illustrated in graphical form in Figure 3. It is observed that the closeness coefficient values are slightly different from each other. A2 has the highest closeness coefficient, which is 0.689, compared to other alternatives. Thus, Apple has ranked in the first place among other smartphone brands evaluated by ten decision makers. A5 follows closely behind A2 with the closeness coefficient value of 0.668. Huawei is ranked as the second brand after Apple, with a closeness coefficient difference of 0.021. On the other hand, the closeness coefficient value of 0.662 belongs to A1, which is nearly tied with A5. With the closeness coefficient difference of 0.006, A1 or Samsung brand is ranked as the third smartphone brand evaluated by decision makers. Next, the ranking continues with A3, which has a closeness coefficient value of 0.646. The Oppo brand is ranked in fourth place because the closeness coefficient value is lower than Samsung brand. Lastly, with the closeness coefficient value of 0.639, A4 or Vivo is ranked as the fifth smartphone brand as evaluated by the decision makers.

Table 12: Ranking of alternatives

Alternatives	Smartphone Brands	Closeness Coefficient	Rank
A1	Samsung	0.662	3
A2	Apple	0.689	1
A3	Oppo	0.646	4
A4	Vivo	0.639	5
A5	Huawei	0.668	2

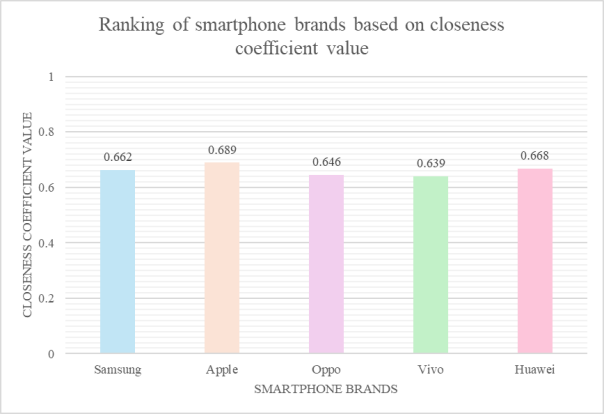


Figure 3: Ranking of smartphone brands based on closeness coefficient value

The findings emphasized the use of MCDM methods in the complex decision-making process. For instance, the use of linguistic variables has significantly helped decision makers to evaluate the criteria and alternatives. By converting the linguistic variables into triangular fuzzy numbers, the subjective opinions from decision makers can be evaluated. Other than that, this method is suitable to be used in

solving complex problems because decision makers are able to evaluate the criteria and alternatives based on their importance. In addition, the method is also straightforward and easy to implement, which makes it understandable even when people do not have extensive knowledge of the method. Furthermore, the implementation of the Fuzzy TOPSIS approach also helps in handling ambiguous and uncertain data. By normalizing the fuzzy decision matrix and applying the fuzzy weight of criteria, the small differences between the values are determined. This makes the approach more reliable to use in ranking some real-world systems. Moreover, this approach has a simpler ranking method compared to other methods, where the alternatives were ranked by using their closeness coefficient value. This approach also helps to minimize inconsistency and bias in the decision-making process because the calculation steps are very detailed and precise. Lastly, this approach promotes balanced assessment as it calculates the distances of alternatives from both positive and negative ideals solutions.

5. Conclusions and Recommendations

Overall, this study aims to apply the Fuzzy TOPSIS approach to the selection of smartphone brands. The first objective of this study is to identify the main criteria that will influence consumers in selecting a smartphone brand before making a purchase on the best smartphone. This study has chosen five main criteria, which are price, RAM, storage, battery and camera, while the alternatives are Samsung, Apple, Oppo, Vivo and Huawei. Since there are too many smartphone brands and models in the marketplace, this study has helped decision makers to focus on several alternatives. To conclude, the first objective is successfully achieved as the main criteria and alternatives are identified.

Furthermore, this study has applied the Fuzzy TOPSIS approach to determine the best smartphone brand among the alternatives listed and using the five main criteria. The Fuzzy TOPSIS approach is an extension of the TOPSIS approach, where this method uses TFN instead of crisp values. Questionnaires are distributed among the committee of ten decision makers and the data obtained was calculated using the Fuzzy TOPSIS approach. The evaluation of decision makers on the weight of criteria and rating of alternatives uses linguistic variables, which are then converted into TFN. The TFN for the weight of criteria and fuzzy decision matrix were obtained by calculating its average. Next, the fuzzy decision matrix was normalized, obtained the weighted normalized fuzzy decision matrix, obtained the FPIS and FNIS and lastly calculated the closeness coefficient values. The closeness coefficient values are used in ranking the smartphone brands.

The smartphone brands were ranked in descending order, starting from A2, followed by A5, A1, A3 and lastly A4. The value that is much closer to 1 is considered the most preferred alternative, while the value that is much closer to 0 is considered the least preferred alternative. This study has found that the most preferred alternative is A2 because it has the highest value of the closeness coefficient, while A4 has the lowest value, thus it was the least preferred alternative. This can be concluded that A2, which stands for Apple brand, is the most preferred alternative among decision makers, followed by A5 or Huawei, A1 which is Samsung, A3 which stands for Oppo and lastly A4 which is Vivo as the least preferred smartphone brand among decision makers.

Based on the findings, there are some recommendations that can be applied to this study for the future. For further extensions of the study, other MCDM methods, such as Analytic Hierarchy Process (AHP) or VIKOR, which can rank both criteria and alternatives can be suggested to be used. This approach will provide a more precise analysis framework based on the future study objectives [1], [15]. Next, other weighting methods that are superior can be suggested to be used to determine the weight of criteria. The suggested weighting methods are subjective methods, which highlight the preferences of decision makers, objective methods that do not consider decision makers' preferences or hybrid methods, which combine both subjective and objective methods [9]. By addressing the suitable recommendations, the study can be extended and the application of the Fuzzy TOPSIS approach and other MCDM methods can be expanded. Therefore, a more precise and flexible output from decision-making problems can be obtained.

Acknowledgement

The authors wish to thank all family members and friends for their invaluable support and assistance during the preparation of this study. Additionally, the authors appreciate the participation of a committee of ten decision makers who offered their time and knowledge to help the authors gather reliable data for this study. Lastly, the authors are grateful to everyone who contributed to making this study a reality, no matter how tiny their contribution was.

References

[1] Abu Bakar, S., Fauzi, N. S., & Abd Rahman, K. (2021). Generalized Fuzzy Number and Graph Theory Matrix Approach in Fuzzy Decision Making of Smartphone Selection. *Journal of Physics: Conference Series, 1988(1),* 1-13. https://iopscience.iop.org/article/10.1088/1742-6596/1988/1/012022/meta

[2] Arifin, M., Sudiati, L. E., & Haryani, N. (2022). Implementation of Technique Method for Order Preference by Similarity to Ideal Solution (TOPSIS) for Smartphone Selection. *SOSCIED, 5(2)*, 238–245. https://doi.org/10.32531/jsoscied.v5i2.528

[3] Chakraborty, J., Mukherjee, S., & Sahoo, L. (2023). Intuitionistic Fuzzy Multi-Index Multi-Criteria Decision-Making for Smartphone Selection Using Similarity Measures in a Fuzzy Environment. *Journal of Industrial Intelligence, 1(1),* 1–7. https://doi.org/10.56578/jii010101

[4] Chen, C. T. (2000). Extensions of The TOPSIS for Group Decision-Making Under Fuzzy Environment. *Fuzzy Sets and Systems, 114(1),* 1–9. https://doi.org/10.1016/s0165-0114(97)00377-1

[5] Goswami, S. S., & Behera, D. K. (2021). Evaluation of The Best Smartphone Model in The Market by Integrating Fuzzy-AHP and PROMETHEE Decision-Making Approach. *DECISION, 48(1),* 71-96. https://doi.org/10.1007/s40622-020-00260-8

[6] Gunawan, F. S., Novelan, M. S., & Arpan. (2022). Application of The TOPSIS Method in Selection of The Best Smartphones at Price of About One Million. *INFOKUM, 10(02)*,1276-1281. http://infor.seaninstitute.org/index.php/infokum/article/view/637

[7] Hossain, R., Hasan, R., & Sharmin, M. (2022). A Short Review on The History of Mobile Phones. *Journal of Android, IOS Development and Testing, 7(2)*, 33-39.

[8] Hsieh, T.-Y., Lu, S.-T., & Tzeng, G.-H. (2004). Fuzzy MCDM Approach for Planning and Design Tenders Selection in Public Office Buildings. *International Journal of Project Management*, *22*(7), 573–584. https://doi.org/10.1016/j.ijproman.2004.01.002

[9] Keshavarz-Ghorabaee, M., Amiri, M., Zavadskas, E. K., Turskis, Z., & Antucheviciene, J. (2021). Determination of Objective Weights Using a New Method Based on the Removal Effects of Criteria (MEREC). *Symmetry*, *13*(4), 1-20. https://doi.org/10.3390/sym13040525

[10] Musrini, M., Umaroh, S., & Arifin, A. A. (2020). Decision Support System Using TOPSIS Method for Smartphone Selection. *Faculty of Industrial Technology International Congress,* 270–276. https://eproceeding.itenas.ac.id/index.php/foitic/article/view/91

[11] Saqlain, M., Jafar, M. N., & Riaz, M. (2020). A New Approach of Neutrosophic Soft Set with Generalized Fuzzy TOPSIS in Application of Smart Phone Selection. *Neutrosophic Sets and Systems, 32,* 307-316.

[12] Shamsuzzoha, A., Piya, S., & Shamsuzzaman, M. (2021). Application of Fuzzy TOPSIS Framework for Selecting Complex Project in a Case Company. *Journal of Global Operations and Strategic Sourcing, 14(3),* 528–566. https://doi.org/10.1108/jgoss-07-2020-0040

[13] Soba, M., Ersoy, Y., Tarakcioğlu Altınay, A., Erkan, B., & Şik, E. (2020). Application of Multiple Criteria Decision-Making Methods in Assignment Place Selection. *Mathematical Problems in Engineering, 20*,1–13. https://doi.org/10.1155/2020/6748342

[14] Surya, A. H., Zufria, I., & Nasution, Y. R. (2021). Decision Support System to Choose Smartphone Using TOPSIS Method. *INFOKUM, 10(1),* 196–205. https://seaninstitute.org/infor/index.php/infokum/article/view/253

[15] Taherdoost, H., & Madanchian, M. (2023). Multi-Criteria Decision Making (MCDM) Methods and Concepts. *Encyclopedia, 3(1)*, 77–87. https://doi.org/10.3390/encyclopedia3010006

Predicting Factors in Financial Loss Among Malaysian Scam Victims Using Machine Learning

Nur Alisa Binti Azian[1], Che Norhalila Binti Che Mohamed[1, *]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawanagn Negeri Sembilan, Kampus Seremban, 70300 Seremban, Negeri Sembilan.*
[*]cheno849@uitm.edu.my

Abstract

This study compares decision tree and logistic regression models to predict financial losses among 394 Malaysian scam victims and to identify key predictors. Model accuracy refers to overall classification accuracy on the validation set, alongside AUC, sensitivity, specificity, and F1 score. Model interpretability was operationalized as the transparency of decision rules in the fitted tree, that is, the ease of tracing split conditions along each decision path. On validation, the decision tree achieved higher AUC and accuracy than logistic regression (AUC 0.838 versus 0.797; accuracy 83.05 percent versus 72.88 percent), with gains in sensitivity and F1 score as well. Emotional harm emerged as the strongest predictor of financial loss, followed by cybersecurity knowledge and age, whereas gender, urbanity, education, and internet use contributed modestly. These findings support targeted victim support, stronger fraud detection, and population-level digital literacy initiatives.

Keywords*:* Decision Tree, Financial Loss, Logistic Regression, Scam Victims, Malaysia

Introduction

The Internet is widely regarded as one of humanity's most remarkable inventions, having permanently transformed communication, information access, and commerce in daily life [1]. In Malaysia, the number of Internet users has continued to rise, from 96.8% in 2021 to 97.4% in 2022 [2]. This increasing dependence on digital platforms and social networking sites has created new opportunities but also heightened exposure to scams and fraudulent activities.

Malaysia has experienced a rapid escalation in online scams, with 95,837 incidents reported between 2021 and April 2024, amounting to financial losses of RM3.18 billion. Reports from the National Scam Response Centre (NSRC) indicate that these figures are likely underestimated, as many victims do not file official complaints [3]. The growing prevalence of scams underscores the urgent need for evidence-based strategies to protect vulnerable groups and minimise financial harm.

This study aims to classify scam victims based on their demographic and knowledge-related characteristics and to identify predictors of financial loss using machine learning techniques. By addressing this gap, the research contributes to understanding the dynamics of scam victimisation in Malaysia and supports the design of more effective preventive measures.

Background

Scams or fraudulent activities, often referred to as confidence tricks, have existed long before the advent of the Internet [4]. Susceptibility to scams is shaped by multiple factors, including the persuasive strategies of perpetrators, cognitive processes of victims, and individual differences in vulnerability [5]. In Malaysia, common scam types include love scams, Macau scams, job scams, parcel scams, and online loan scams [6].

Prior research has examined demographic and psychological risk factors associated with scam victimisation. However, limited attention has been devoted to applying advanced analytical methods to predict financial loss. Machine learning techniques, such as decision trees and logistic regression, are well-suited to this task because they can manage complex interactions among variables and generate accurate predictions [7,8]. Decision trees, in particular, provide intuitive visualisations that enhance interpretability for policymakers and stakeholders [9]. Recent studies highlight their potential in identifying how demographic, behavioural, and knowledge-based factors contribute to vulnerability [10].

Despite this progress, research employing machine learning for scam prediction in Malaysia remains scarce. Machine learning, as a branch of artificial intelligence (AI), offers robust tools capable of detecting subtle patterns those traditional statistical methods may overlook [11]. Leveraging these approaches may generate novel insights into predictors of financial loss and inform targeted interventions to reduce the burden of scams.

Methodology

Research Design

This study used both descriptive and exploratory research designs. A descriptive design is used to describe the type of scam experienced by the victims to find out the amount of losses incurred by the victims of scams in Malaysia. An exploratory design tests the hypotheses based on the theoretical framework. A cross-sectional design was implemented in this study because it captures data at a single point in time, and the conclusions drawn are valid only during the period in which the research was conducted.

Population and Sample

The study targeted Malaysians who had experienced scams, irrespective of the type of scam. Due to the unknown population of scam victims, the minimum sample size for categorical data was computed using Cochran's formula with $p = 0.50$ and a 5% margin of error, which yielded 384 respondents at a 95% confidence level. This decision also falls within the recommended sample range of 30 to 500 for behavioral research [12], supporting its suitability for the present study.

Sampling Method

Sampling methods are categorised into probability sampling, where every individual has an equal chance of selection, and non-probability sampling, which relies on subjective judgment [13]. Due to the confidentiality of scam victims' data from the Royal Malaysia Police (Polis Diraja Malaysia, PDRM), no sampling frame was available, necessitating the use of non-probability sampling. Convenience sampling was chosen, relying on accessible primary data sources for faster, cost-effective data collection. However, convenience sampling is inherently prone to bias because it does not ensure that the sample is representative of the entire population, as it selects participants based on their availability and willingness to participate [14,15]. This method can overrepresent certain groups and underrepresent others, leading to results that may not accurately reflect the broader population. Consequently, findings are generalisable only to the sampled subpopulation, limiting their applicability to broader contexts [16,17].

Data Collection Method

Primary data were collected directly from scam victims using a quantitative approach. An online survey was administered through Google Forms and distributed via multiple social media platforms, including Facebook, WhatsApp, Telegram, and Instagram. The recruitment period lasted two months after obtaining ethical approval, during which the questionnaire links were actively disseminated to family members, friends, and members of scam victims' groups on Facebook to maximise participation. This strategy enabled the collection of firsthand information on demographic characteristics, internet safety knowledge, and the financial, emotional, and behavioural impacts of scams. Online surveys were selected due to their speed, accuracy, cost-effectiveness, and flexibility, making them well-suited for this research [18].

Research Instrument

The research instrument used in this study was a 29-question questionnaire divided into three parts. Part A included demographic questions such as gender, age, education level, internet usage, urbanity, and scam type. Part B consisted of 12 questions on internet safety knowledge, based on the Knowledge and Attitude Towards Internet Safety Questionnaire (KATISQ) [19], and three additional questions about the scamming channels, reporting behaviour, and recipients of reports [20]. Part C focused on the negative impacts of scams, covering financial, emotional, and behavioural effects, including financial losses, with seven questions in total [20]. Overall, the designed questionnaire, with detailed descriptions of the variables provided in Table 1.

Table 1: Questionnaire Item.

Part	Item	Number of Questions	Question Number
A	Demographic	6	1 – 6
B	Knowledge about internet safety	12	7 - 18
	Scammed channel	1	19
	Reported scam	1	20
	Reported a scam to	1	21
C	Negative impact	6	22 - 28
	Negative attitude	1	29

Theoretical Framework

Figure 1 illustrates the research theoretical framework of this study, in which the independent variables are demographic factors (age, gender, level of education, internet usage and urbanity), knowledge about internet safety and negative impacts (emotional impact and physical impact) are the factors that influence the financial loss experienced by scam victims.

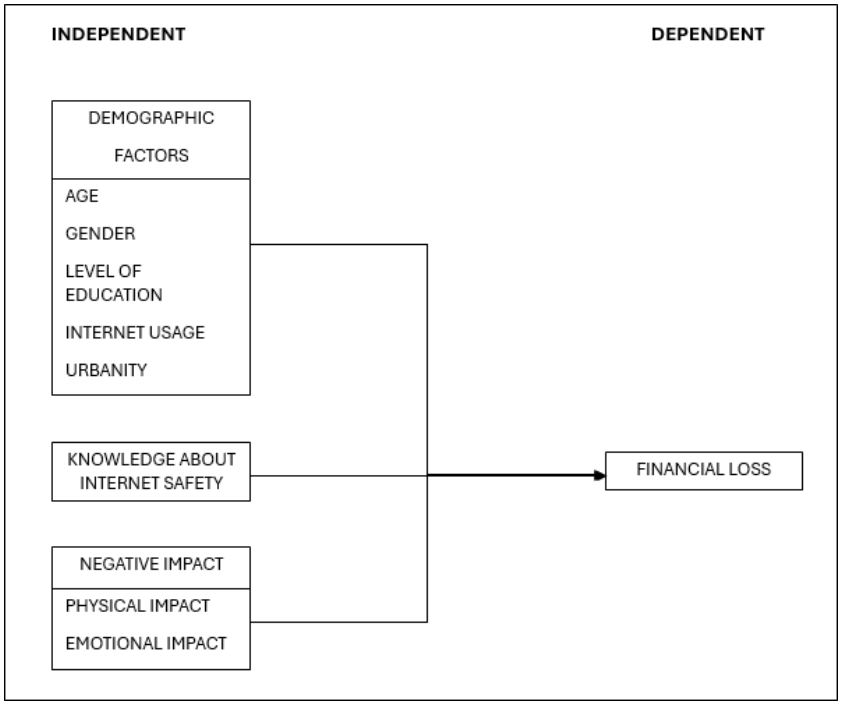


Figure1: Research Theoretical Framework.

Data Analysis

Descriptive analysis was employed to summarise the traits of individuals who have fallen victim to scams. By identifying shared characteristics and vulnerabilities among scam victims, this analysis provides insights into those most at risk and informs tailored interventions to protect them. This examination highlights the demographic trends of scam victimisation in Malaysia, supporting a more effective strategy to combat scams and mitigate their impact on individuals and society. Additionally, the analysis focused on identifying the best predictive model and the factors influencing financial losses among scam victims. Logistic regression and decision tree analysis were utilised, with model performance evaluated using metrics such as classification accuracy, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC) to identify the factors influencing the financial losses experienced by scam victims in Malaysia. Logistic regression highlighted significant predictors ($p < 0.05$), while decision tree analysis ranked variables based on importance scores. This approach provided a comprehensive understanding of the key factors contributing to financial losses.

Machine Learning Models and Evaluation

Two supervised learning algorithms were trained to classify financial loss status, namely logistic regression and a Classification and Regression Tree (CART)-style decision tree using Gini impurity as the splitting criterion. Logistic regression estimates the log-odds of loss as a linear function of predictors and yields adjusted odds ratios for inference and communication [21]. In contrast, decision trees iteratively partition the predictor space into a series of decision rules that can be directly inspected, offering greater transparency and interpretability for stakeholders and policymakers [22]. Model performance was evaluated on the validation dataset using several complementary measures: the area under the receiver operating characteristic curve (AUC), overall classification accuracy, sensitivity, specificity, and the F1 score [23]. Interpretability of the decision tree was operationalized in terms of rule transparency, defined as the ease of tracing conditions from root to leaf, which provides an intuitive supplement to quantitative performance metrics. Variable importance values were also extracted from the tree to rank predictors according to their relative contribution to classification.

Result and Discussion

Descriptive Analysis

The demographic findings in Table 2 show that females are slightly more susceptible to scams, though the difference is not significant. Older individuals aged 55 and above are more likely to experience financial loss, while those with only primary or secondary education are particularly vulnerable. This highlights the critical need for financial literacy and scam awareness programs to equip these groups better. Moderate internet users are at a higher risk of financial loss compared to light or heavy users, possibly due to inconsistent online habits. Urban residents are also more prone to scams, likely because of greater exposure and higher levels of online activity. These findings underscore the importance of developing tailored, urban-focused scam

prevention strategies and targeted educational initiatives to reduce vulnerabilities across these demographic groups in Malaysia.

Table 2: Summary of Demographics of Scam Victims

Variable		Financial Loss (Yes)		Financial Loss (No)	
		N	%	N	%
Gender					
	Female	129	32.7	107	27.2
	Male	104	26.4	54	13.7
Age (years)					
	18–24	100	25.4	100	25.4
	25–34	51	12.9	22	5.6
	35–54	62	15.7	35	8.9
	≥55	20	5.1	4	1
Education level					
	Primary School	1	0.3	0	0
	High School	28	7.1	16	4.1
	Diploma	87	22.1	45	11.4
	Bachelor	103	26.1	89	22.6
	Postgraduate	14	3.6	11	2.8
Internet Usage					
	Light	35	8.9	17	4.3
	Moderate	143	36.3	102	25.9
	Heavy	55	14	42	10.7
Urbanity					
	Urban	136	34.5	89	22.6
	Suburban	61	15.5	47	11.9
	Rural	36	9.1	25	6.3
Total		233	59.1	161	40.9

Based on the Figure 2, online sales scams stand out as the most common, with a significant number of victims experiencing financial loss. Short Message Service (SMS) scams also demonstrate a concerning prevalence, followed by Macau scams and job scams that contribute notably to financial loss. Additionally, the chart also reveals that African scams and business email compromise have relatively fewer victims but are still areas of concern. The findings highlight the necessity for targeted interventions and educational programs tailored to specific scam types. The findings highlight that online sales and SMS scams are the most prevalent, posing significant risks due to their digital and mobile nature. Macau and job scams also

contribute notably to financial losses, while African scams and business email compromise remain areas of concern.

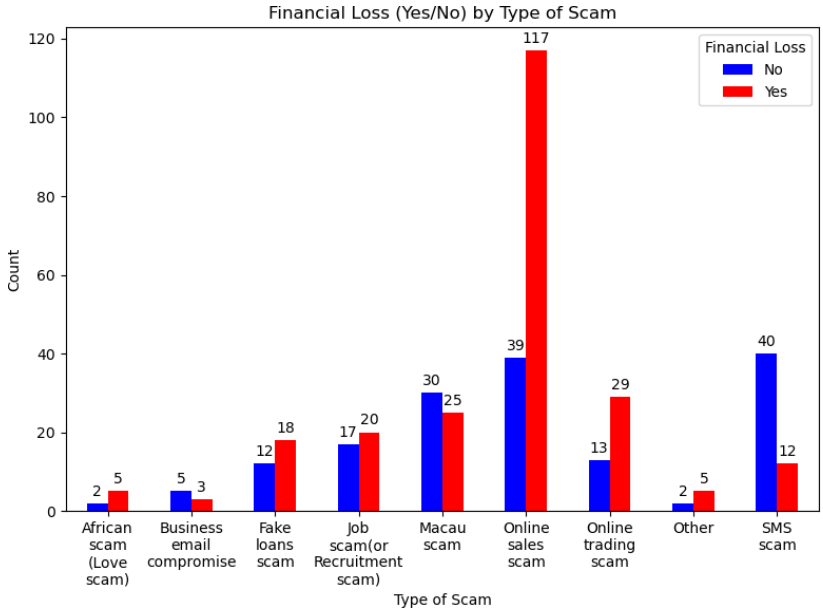


Figure 2: The Distribution of Scam Types by Financial Loss (Yes/No).

Best Model for Predicting the Financial Losses Experienced by The Scam Victims in Malaysia

According to Table 3, the decision tree model outperformed logistic regression in predicting financial loss, as evidenced by higher AUC, accuracy, sensitivity and F1 score. Firstly, the decision tree model demonstrated a higher AUC value on the validation dataset (AUC = 0.838) compared to the logistic regression model (AUC= 0.797). This indicates that the decision tree model generalises better to unseen data and is more likely to predict financial loss in real-world scenarios accurately. Furthermore, the decision tree model outperformed the logistic regression model in terms of accuracy (83.05% vs. 72.88%) and sensitivity (87.14% vs. 72.86%), highlighting its effectiveness in correctly identifying individuals who experienced financial loss. While the logistic regression model showed slightly higher specificity (72.92% vs. 77.08%), the overall performance of the decision tree model, as reflected by the F1 score (85.92% vs. 76.12%), suggests its superiority in predicting financial loss among scam victims. These insights reinforce the decision tree model's potential as a valuable method for mitigating financial loss among the scam victims in Malaysia.

Table 3: Model Performance Comparison

Comparison Metrics	Decision Tree (GINI)	Logistic Regression (Main Effect)
AUC	0.838	0.797
Accuracy	83.05%	72.88%
Sensitivity (Recall)	87.14%	72.86&
Specificity	72.92%	77.08%
F1 Score	85.92%	76.12%

Factors That Influence the Financial Losses Experienced by The Scam Victims in Malaysia

Given that the decision tree model outperformed the logistic regression model in predictive accuracy and reliability, Table 4 summarizes the key results of the decision tree analysis. Emotional harm was identified as the most influential variable with an importance score of 1.0000, indicating its critical role in distinguishing between victims who experienced financial losses and those who did not. Knowledge score was the second most important factor (0.5123), suggesting that individuals with higher internet knowledge were better classified by the model. Age ranked third (0.3947), reflecting a moderate influence likely tied to variations in exposure or susceptibility across age groups. Gender contributed less significantly (0.1730), while urbanity (0.1152), education level (0.0947), and internet usage (0.0821) showed limited predictive power. Physical harm had an importance score of 0.0000, indicating no contribution to the classification. These results suggest that emotional harm is the primary predictor, followed by knowledge and age, while other factors play minor roles.

The predominance of emotional harm as the strongest predictor aligns with growing evidence that psychological distress plays a central role in shaping victims' financial decision-making under fraudulent influence. Victims who experience heightened emotional states such as fear, urgency, or shame are more vulnerable to persuasive tactics and more likely to comply with scammers' demands, thereby increasing the probability and magnitude of financial loss. Prior research has shown that the severity of emotional harm is strongly correlated with financial vulnerability: identity theft victims reporting higher distress tend to incur larger losses [24], while more recent qualitative studies of romance scams demonstrate that emotional manipulation intensifies susceptibility to substantial monetary harm [25,26]. These findings suggest that emotional harm is not merely a consequence of fraud but functions as a key mechanism of victimization, underscoring the need for integrated psychosocial support in addition to financial protection interventions.

Table 4: Feature Importance.

Variable	Importance Score
Emotional harm	1.0000
Knowledge score	0.5123
Age	0.3947
Gender	0.1730
Urbanity	0.1152
Level of education	0.0947
Internet usage	0.0821
Physical harm	0.0000

Conclusion

The main aim of this study was to determine the best predictive model between the decision tree and logistic regression and to identify the factors influencing financial losses among scam victims in Malaysia. The findings revealed that the decision tree model outperformed logistic regression in predictive accuracy and classification performance, making it the superior choice for this analysis. The study highlighted emotional harm as the most significant predictor of financial loss, followed by internet safety knowledge and age, while gender, urbanity, education, and internet usage played minor roles. Notably, physical harm showed no influence on financial losses.

Addressing emotional harm and improving digital literacy are essential strategies for mitigating financial losses. These findings underline the need for targeted public education campaigns, robust fraud detection through collaboration between the government and financial institutions, and expanded victim support services.

Future research should further extend this work by integrating advanced machine learning techniques, such as neural networks and ensemble learning models, to capture more complex and nonlinear relationships between predictors of financial loss. Additionally, longitudinal datasets on scam victimization would enable the examination of temporal dynamics, such as changes in vulnerability and recovery patterns over time, which cannot be captured in cross-sectional studies. Incorporating these approaches could provide a deeper understanding of the causal mechanisms underlying scam victimization and improve the predictive validity of prevention models.

Acknowledgements

The authors would like to express their sincere gratitude to all respondents who generously shared their time and experiences, without whom this study would not have been possible. Appreciation is also extended to Universiti Teknologi MARA (UiTM) and the College of Computing, Informatics, and Mathematics for providing the necessary resources and support.

The authors are deeply thankful to the supervisor for her invaluable guidance and encouragement throughout the study, and to colleagues and lecturers for their constructive input. Finally, heartfelt thanks are given to family and friends for their continuous support and motivation.

References

[1] Brooks, R. (2022). *The invention and evolution of the Internet*. Wrexham University Online. https://online.wrexham.ac.uk/the-invention-and-evolution-of-the-internet

[2] Department of Statistics Malaysia. (2022). *ICT use and access by individuals and households survey report, Malaysia*. Putrajaya: DOSM.

[3] Mahaizura, A. M. (2024, August 1). *Kes tipu direkod paling tinggi dalam jenayah atas talian*. Harian Metro. Retrieved September 9, 2024, from https://www.hmetro.com.my/mutakhir/2024/08/1121860/kes-tipu-direkod-paling-tinggi-dalam-jenayah-atas-talian

[4] Whitty, M. T. (2020). Is there a scam for everyone? Psychologically profiling cyberscam victims. *European Journal on Criminal Policy and Research, 26*(3), 399–409. https://doi.org/10.1007/s10610-019-09425-9

[5] Hanoch, Y., & Wood, S. (2021). The scams among us: Who falls prey and why. *Current Directions in Psychological Science, 30*(3), 260–266. https://doi.org/10.1177/09637214211043460

[6] Majlis Keselamatan Negara. (2023, March 28). *Jenis penipuan atas talian*. Laman Web Rasmi MKN. Retrieved August 27, 2024, from https://www.mkn.gov.my/web/ms/2023/03/28/jenis-penipuan-atas-talian/

[7] Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST, 25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

[8] Zhang, Y., Chen, J., & Hu, Y. (2021). Exploring machine learning methods for financial fraud detection: A review. *Financial Innovation, 7*(1), 1–20. https://doi.org/10.1186/s40854-021-00232-9

[9] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.

[10] Nguyen, T., Le, H. D., & Tran, Q. (2023). Decision tree models in cybersecurity risk analysis: Applications and future directions. *Journal of Cybersecurity Research, 12*(2), 56–73.

[11] Brown, S. (2021). *Machine learning, explained*. *MIT Sloan School of Management*. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

[12] Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill-building approach* (7th ed.). Wiley.

[13] Bhardwaj, P. (2019). Types of sampling in research. *Journal of the Practice of Cardiovascular Sciences, 5*(3), 157–163. https://doi.org/10.4103/jpcs.jpcs_62_19

[14] Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review, 33*(4), 357–370. https://doi.org/10.1016/j.dr.2013.08.003

[15] Sedgwick, P. (2013). Convenience sampling. *BMJ, 347*, f6304. https://doi.org/10.1136/bmj.f6304

[16] Andrade, C. (2021). The inconvenient truth about convenience and purposive samples. *Indian Journal of Psychological Medicine, 43*(1), 86–88. https://doi.org/10.1177/0253717620977000

[17] Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). More than just convenient: The scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development, 82*(2), 13–30. https://doi.org/10.1111/mono.12296

[18] Salkind, N. J. (2010). *Encyclopedia of research design* (Vol. 1). SAGE.

[19] Okanlawon, A. E., Yusuf, F. A., & Abanikannda, M. O. (2015). University students' knowledge and attitude towards Internet safety: A preliminary study. *Journal of Emerging Trends in Educational Research and Policy Studies, 6*(3), 279–286.

[20] Bijwaard, D. (2020). *Survey on "scams and fraud experienced by consumers" — Final report*. European Institute for Gender Equality. Lithuania. Retrieved January 15, 2025, from https://commission.europa.eu/system/files/2020-01/survey_on_scams_and_fraud_experienced_by_consumers_-_final_report.pdf

[21] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

[22] Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: From efficient prediction to responsible AI. *Frontiers in Artificial Intelligence, 6*, 1124553. https://doi.org/10.3389/frai.2023.1124553

[23] Çorbacıoğlu, Ş. K., & Delil, Ş. (2023). A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine, 23*(4), 165–169. https://doi.org/10.4103/tjem.tjem_155_23

[24] DeLiema, M., Shadel, D., Pak, K., & Niemiec, R. (2021). The financial and psychological impact of identity theft. *Frontiers in Psychology, 12*, 777043. https://doi.org/10.3389/fpsyg.2021.777043

[25] Cole, R., Anderson, S., & Ward, C. (2024). A qualitative investigation of the emotional, physiological, and financial impacts of online romance scams. *Social Science & Humanities Open, 10*(1), 100060. https://doi.org/10.1016/j.ssaho.2024.100060

[26] Wang, J., Liu, X., & Li, H. (2024). The dynamic emotional experience of online fraud victims. *Journal of Criminal Justice, 90*, 102112. https://doi.org/10.1016/j.jcrimjus.2024.102112

Prediction Model of Home Loan Eligibility Status

Danisya Adlina Ruzi[1], Siti Aisyah Mohamed Kher[1], Isnewati Ab Malek[1,], Haslinda Ab Malek[1*,]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia*
[*]Corresponding: haslinda8311@uitm.edu.my

Abstract

A home loan is known as a mortgage which is a type of loan used to purchase a home or real estate. Home loans are essential for many individuals who are looking to buy a home without paying the entire price upfront. Confronting a prevalent challenge involving a notable of 60% rejection rate among home loan applicants, this issue was identified during the pre-pandemic year. Therefore, this study aims to develop an optimized home loan eligibility status prediction model. The study systematically evaluated Decision Tree approaches, incorporating variable selection for enhanced precision in model comparison. The findings of this study found that loan amount, marital status, property area, and co-applicant income as pivotal factors influencing home loan eligibility status. Notably, the Decision Tree Entropy model emerges as the optimal model, achieving a remarkable average squared error of 19.71%. This model exhibits superior performance, evidenced by a high sensitivity of 84.16% and accuracy of 59.78% according to the confusion matrix. The results are hoped to provide valuable insights for improving house loan eligibility prediction models, which could improve the banking sector decision-making procedures and in turn, improve consumer satisfaction.

Keywords*:* Housing loan, Loan Approval, Decision Tree.

Introduction

A home loan is a loan that helps a borrower purchase a home. The amount borrowed from a home is based on the appraised value of the home and the amount of money paid as a down payment. The home will be put up as collateral and loan applicants will make monthly mortgage payments for a specified amount of time (typically 15 or 30 years) until the loan is paid off. According to a study done by [1], in Malaysia, home loans can be divided into three main categories: Basic Term Loans, Semi-Flexi Loans, and Full-Flexi Loans.

In this study, the home loan eligibility status is predicted. The term 'home loan eligibility status' defines an applicant's suitability for a home loan based on the requirements set by the lender. It indicates if the applicants meet the requirements for a home loan based on variables such as income and loan amount. Home loan eligibility status is an important factor in the lending industry, and financial institutions must predict the likelihood of loan approval accurately. However, according to the finding by [2], an average of 60% of the home loan applicants failed to get a home loan during the pre-pandemic year. This has been a problem and a significant issue for the home loan applicants.

Therefore, this study was conducted to construct the home loan eligibility status model in determining the type of model that is best to use. The model can help to analyse applicant data more rapidly and improve the overall customer experience by offering faster and more accurate predictions. The application of these models can assist banks and financial institutions to make more informed home loan approval choices. Moreover, these models can reduce the risk of loan default and improve the overall efficiency of the home loan approval process.

As the demand for home loans in banks grows by the day, [3] performed a Modern Approach to Home Loan Sanctioning in Banks Using Machine Learning. The goal is to apply machine learning tools that use classification algorithms to predict worthy applicants for home loan approval as an appropriate option to reduce human efforts and make successful decisions in the home loan approval process. A system to develop a model is built in their research by training the system with records and approval results of previously applied home loan applicants

Moreover, study done by [4] employed two machine learning classification models, namely Decision Tree and Random Forest, for loan prediction. The three findings indicate that the highest level of accuracy achieved is

62.12%, accompanied with a confidence factor of 0.15. Other than that, [5] employed decision trees to construct an accurate and efficient predictive model for forecasting home loan approval. Experiments were conducted utilising several tree methodologies, spanning from the simplest and most comprehensible decision tree to the most intricate random forests.

Hence, [6] did a comprehensive study to make a detailed comparison between the Random Forest and Decision Trees algorithms. Both methodologies were employed on the identical dataset, and the outcomes revealed that the Random Forest algorithm exhibited much superior performance in terms of accuracy compared to the Decision Tree strategy, achieving a score of 80% as opposed to 73%. Moreover, [7] conducted a study to forecast the eligibility of consumers for home loan approval. Three machine learning models employed were the Logistic Regression Model, the Decision Tree, and the Random Forest. The study revealed that based on the comparison of the three models, Logistic Regression exhibits the highest accuracy rate of 89.7059%, followed by Decision Tree with an accuracy rate of 85.4054% and Random Forest with an accuracy rate of 77.4566%.

Based on the feature analysis run by [8], it showed that those who have a higher credit history have a better probability of getting a loan compared to those who do not. Furthermore, [9] conducted a study stated that the home loan approval status is determined to be independent of the applicant's income, gender, and loan amount. However, it appears that the credit history has the greatest impact on loan acceptance, suggesting that an applicant's credit history is used as an initial filter for loan applications.

The findings of this study will provide banking institutions with the best model for determining eligibility for housing loans. The greater demand for property loans justifies the need for a more effective, speedy, and prompt approach for bankers in predicting potential candidates. For the researchers, the study will help them uncover the best model for predicting the loan approval process that many researchers were not able to explore. Moreover, this study not only helps the applicant but also helps the bank institution by minimizing the risk and reducing the number of defaulters.

Methodology

Source of data

This study employed secondary data consisted of 11 variables and 443 observations, with the Loan Status attribute being the variable of interest. The dataset was obtained from the Analytics Vidhya Loan Prediction competition, which provides real-world loan application data for predictive modeling [10]. The variables analysed in this study were displayed in Table 1.

Table 1: Description of the Variables

No	Variable	Type of Variables	Scale	Description
1	Gender	Categorical	Nominal	Male, Female
2	Marital Status	Categorical	Nominal	Yes, No
3	Number of Dependents	Continuous	Interval	0, 1, 2, 3+
4	Education Status	Categorical	Nominal	Graduate, Not Graduate
5	Self-Employment	Categorical	Nominal	Yes, No
6	Applicant's Income	Continuous	Interval	$0 - $22000
7	Co-Applicant's Income	Continuous	Interval	$0 - $11000
8	Loan Amount	Continuous	Interval	$15000 - $375000
9	Loan Term	Continuous	Interval	126.68 - 480 (months)
10	Property Area	Categorical	Nominal	Urban, Semiurban, Rural
11	Loan Status	Categorical	Nominal	Yes, No

Decision Tree

The purpose of employing a Decision Tree in this study is to develop a model that can be used to predict the class or value of the home loan eligibility status by learning simple decision rules inferred from prior data. This study employed three of the most common types of Decision Trees, namely Gini, CHAID, and Entropy. These types are distinguished from each other by the mathematical model used to select splitting property when extracting Decision Tree rules [11]

Model Assessment

Accuracy measures can be used to evaluate the ability of the estimated model to predict the target. The Receiver Operating Characteristic (ROC) Chart, Confusion Matrix, Misclassification Rate, Average Squared Error (ASE), and Receiver Operating Characteristic (ROC) Index are the tools from which different accuracy measures are derived.

Misclassification Rate

The misclassification rate is the frequency with which a classification model guesses a sample's class or category in a dataset incorrectly. It serves as an indicator of how well the model classifies the data. When assessing a classification model, the misclassification rate is determined by comparing the sample true class labels with the predicted class labels. The misclassification rate, for example, would be 10% if a model properly classified 90 out of 100 samples, meaning that 10% of the data were misclassified.

Average Squared Error

The average squared error (ASE) measures how close a regression line comes to a set of data points. The model with a smaller value of ASE indicates a better fit. Hence, it is better suited to be used in predicting the target. Equation 1, as shown, can be used to calculate the average squared error

$$\mathrm{ASE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{1}$$

Receiver Operating Characteristic (ROC) Index

The Receiver Operating Characteristic (ROC) index is the area under the value (AUC) in the ROC chart. The AUC is widely used to measure the accuracy of diagnostic tests. The higher the value of the ROC index, the greater the sensitivity and accuracy of the model.

Confusion Matrix

Accuracy is measured regarding the error rate percentage of records classified correctly or incorrectly. A confusion matrix can be used to determine this measurement. The number of accurate and inaccurate predictions made by the classification model concerning the actual target variable in the data is displayed in a confusion matrix.

Results and Discussion

Model comparisons between Decision Tree (DT) Gini, Decision Tree (DT) Entropy and Decision Tree (DT) CHAID were computed to compare the training sample and validation sample based on the values of the misclassification rate, average squared error, and Receiver Operating Characteristic (ROC) index.

Table 2: Model Comparison Based on Misclassification Rate, Average Squared Error and ROC Index

Model	Misclassification Rate			Average Squared Error			ROC Index		
	Valid	Train	Gap	Valid	Train	Gap	Valid	Train	Gap
DT Gini	0.430	0.307	0.123	0.278	0.197	0.081	0.494	0.735	-0.241
DT Entropy	0.413	0.303	0.110	0.291	0.197	0.094	0.570	0.722	-0.152
DT CHAID	0.458	0.386	0.072	0.247	0.232	0.015	0.559	0.621	-0.062

Table 2 shows the values of valid, train, and gap for the misclassification rate, average squared error, and ROC index for each model. The valid and gap values are used to find the best model. The gap values are calculated by finding the difference between valid and train values. The negative sign is not taken into consideration for the gap value. From Table 2, it is found that the DT Gini has the highest gap value for the misclassification rate (0.123) and ROC index (0.241). Therefore, DT Gini is found to be the overfit model.

Additionally, the valid value is evaluated to determine the best fit model. The best model has the lowest valid values for the misclassification rate and average squared error. In contrast, the model must have the highest valid value of the ROC index. DT Entropy has the lowest valid misclassification rate (0.413) and the highest ROC index valid value (0.570). Hence, DT Entropy is concluded as the best model since it meets most of the conditions.

Table 3 summarise that there is no underfit model since all model performances have the best validation results and no negative valid value. However, there is one overfit model, which belongs to the Decision Tree Gini. An overfit model can be determined when it has the largest gap value between valid and train value. Thus, the Decision Tree Gini is an overfitted model because it has the largest gap values for the misclassification rate and the ROC index. In choosing the best models, the valid values for the misclassification rate and average squared error need to be the lowest, while the value for the ROC index needs to be the highest. The Decision Tree Entropy meets most of the requirements, hence, it is chosen as the best model.

Table 3: Summary of Model Comparison

Assesment of Generalization	**Model Description**
Underfit	None
Overfit	Decision Tree Gini
Best Fit	Decision Tree Entropy

Best Model

The confusion matrix is observed to further confirm that the Decision Tree Entropy is the best model. A confusion matrix is a table that describes the performance of a classification model on a set of test data. The factors in the confusion matrix can be examined to yield several performance measures such as sensitivity, specificity, and accuracy, which provide insights into the model's performance.

Table 4: Result Analysis of Confusion Matrix

Model	**Sensitivity**	**Specificity**	**Accuracy**
Decision Tree Gini	0.6931	0.4231	0.5754
Decision Tree Entropy	0.8416	0.2821	0.5978
Decision Tree CHAID	0.4752	0.6281	0.5419

Based on Table 4, the Decision Tree CHAID has the highest specificity values at 62.81%, showing that the model is the best model for classifying the 'No' group of home loan eligibility status. Meanwhile, the Decision Tree Entropy

has the highest sensitivity and accuracy values at 84.16% and 59.78%, respectively. This confirmed that Decision Tree Entropy is the best model for classifying the 'Yes' group of home loan eligibility status and the best in predicting the overall performance of the model.

Conclusion

The specific requirements for obtaining a home can be ambiguous, causing applicants to question the qualifying conditions and the procedures involved in applying. Thus, this study has been conducted using secondary data obtained which includes ten variables that are believed to have an impact on the approval of house loans. This dataset has 443 observations with a binary target variable. The Decision Tree Gini is found to be overfit while none of the models are underfit. It was discovered that implementing Decision Tree Entropy produces the best model for predicting a house loan applicant's eligibility status with an average squared error of only 19.71% and highest accuracy values of 59.78%. Consequently, the Decision Tree Entropy model was deemed to be the most reliable in predicting home loan eligibility status. It is hoped that this research provides valuable insights for home-loan applicants, banking institutions and other researchers for their own interest. This study is in line with a study done [12], it indicates the entropy reduction decision tree technique outperforms the other seven methods in terms of overall classification accuracy rates at the 0.5 cut-off.

For future studies, it is recommended explore other models apart from DT Gini, Decision Tree Entropy and Decision Tree CHAID. One example of an alternative decision tree model is the Random Forest algorithm. Random Forest combines multiple decision trees to make predictions, offering improved accuracy and reducing the risk of overfitting. By utilising a variety of models, researchers can evaluate their performance and compare the results to identify the most effective model for predicting home loan eligibility status.

References

[1] Hassan, M. M., Ahmad, N., & Hashim, A. H. (2022). Housing Property Investment Opportunity in Malaysia: Things that new investor should know. International Journal of Academic Research in Business and Social Sciences, 12(1), 924–941. https://doi.org/10.6007/IJARBSS/v12-i1/12003.

[2] Sooi, C. C. (2021, Sep). Mitigating malaysia's worrying trend of high home loan rejection focus malaysia. Retrieved from https:// focusmalaysia.my/mitigating-malaysias-worrying-trend-of-high-home-loan-rejection.

[3] Rath, G. B., Das, D., & Acharya, B. (2021). Modern approach for loan sanctioning in banks using machine learning. In Advances in machine learning and computational intelligence: Proceedings of icmlci 2019 (pp. 179–188).

[4] Gautam, et al. (2020). Loan prediction using decision tree and random forest. International Research Journal of Engineering and Technology (IRJET), 7(08), 853–856.

[5] Alaradi, M., & Hilal, S. (2020). Tree-based methods for loan approval. In 2020 international conference on data analytics for business and industry: Way towards a sustainable economy (icdabi) (pp. 1–6).

[6] Madaan, et al. (2021). Loan default prediction using decision trees and random forest: A comparative study. In Iop conference series: Materials science and engineering (Vol. 1022, p. 012042).

[7] Dutta, P. (2021). A study on machine learning algorithm for enhancement of loan prediction. International Research Journal of Modernization in Engineering Technology and Science, 3.

[8] Sunitha, T., Chandravallika, M., Ranganayak, M., Suma Sri, G., Jagadeesh, T. V. S., & Tejaswi, A. (2021). Predicting the Loan Status using Logistic Regression and Binary Tree. Paper presented at ICICNIS 2020. http://dx.doi.org/10.2139/ssrn.3769854.

[9] Alaradi, M., & Hilal, S. (2020). Tree-based methods for loan approval. In 2020 international conference on data analytics for business and industry: Way towards a sustainable economy (icdabi) (pp. 1–6).

[10] Analytics Vidhya. (2018). Loan prediction practice problem: Using past data to predict loan eligibility. Retrieved from https://datahack.analyticsvidhya.com/contest/practice-problem-loan-prediction-iii

[11] Frempong, N. K., Nicholas, N., & Boateng, M. (2017). Decision tree as a predictive modeling tool for auto insurance claims. International Journal of Statistics and Applications, 7(2), 117–120

[12] Zurada, J. (2010). Could decision trees improve the classification accuracy and interpretability of loan granting decisions? In 2010 43rd hawaii international conference on system sciences (p. 1-9). doi: 10.1109/HICSS.201



An Analysis of Factors Affecting Automobiles Resale Value in Malaysia

Syahmina Humaira Binti Hizir[1], Noorezatty Binti Mohd Yusop[1,*]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban, Malaysia*
[*]noorezatty@uitm.edu.my

Abstract

Used car buyers often face challenges in understanding the factors that influence the resale value of cars. Identifying these key factors can be difficult, as many elements, such as brand, model, age, mileage, condition, and maintenance history, affect how much a used car is worth. Without understanding the factors that impact used car values, buyers may risk overpaying, especially for those unfamiliar with the used car market. This study examines the resale value of used cars in Malaysia, focusing on the Perodua brand, by analyzing how transmission types, car age, mileage, engine capacity, and body types affect resale value. Secondary data from the Carro website was utilized for this study, covering the period from July 2024 to October 2024. The dataset consisted of 156 observations, focusing only on Perodua cars. Multiple Linear Regression (MLR) was used to identify significant factors affecting the resale value of Perodua cars. All assumptions of Multiple Linear Regression (MLR) were met to ensure the validity of the analysis, which were linearity, normality, homoscedasticity, absence of outliers, multicollinearity and autocorrelation. The regression results concluded that transmission types, mileage (km), age of car (month), engine capacity, and body types were significant factors of the resale value. Recommendations for future studies include extending the data collection period, extracting data from multiple sources, and exploring additional factors.

Keywords*:* Perodua, Multiple Linear Regression, Resale value

Introduction

Presently, most individuals rely on personal modes of transportation, such as cars, with emphasis on the latest models, often without adequate knowledge about used cars. The term ‘used cars’ or ‘resale cars’ refers to cars previously owned by one or more individuals who intend to sell them [1]. In addition, used cars undergo repairs and upgrades involving mechanical issues, thorough cleaning of both interior and exterior and replacing worn-out components to enhance the car’s quality to resell the car to another owner [2]. In this study, the term automobile was also called car, and the term resale value was used to represent the asking price of a used car.

The used car market thrives by offering consumers affordability, a wider range of models, and lower depreciation costs compared to new cars. It appeals to budget-conscious buyers, first-time car owners, and those seeking specific older models or features. Used cars also avoid the

steep initial depreciation of new vehicles, making them a practical option for many. According to Mordor Intelligence, Malaysia's used car market is projected to grow from RM105.65 billion in 2024 to RM148.08 billion by 2029, with a compound annual growth rate (CAGR) of 5.40%. This highlights the growing importance and potential of the sector.

Other factors influencing car purchases include the car's attributes like type, color, model, transmission (automatic or manual), and registration year [1]. Buyers often select MPVs for larger families, while sedans and hatchbacks suit single individuals or small families. SUVs, coupes, and convertibles are also popular choices depending on individual preferences.

For this research, the primary objective is to determine the key factors that significantly influence the car's values. This research was crucial because understanding such factors can provide beneficial insights for stakeholders. Recognising which factors affect resale value helps consumers to make informed choices about buying and maintaining cars, potentially increasing their return on investment. Multiple Linear Regression is appropriate for analysing the multiple variables that impact car resale values, including types of car, age, condition, mileage, and other factors. Past research has employed comparable methodologies to explore these factors, but often adopted a broader scope, encompassing multiple brands or regions [3–5]. This study seeks to fill this gap by concentrating specifically on Perodua, Malaysia's car brand, aiming to offer targeted and practical insights that are directly applicable to the Malaysian automotive market.

Factors of Resale Value

Transmission types

Car transmission is among the most crucial elements of a car where they are responsible for transferring engine power to the wheels [6]. Car's transmission comes in two types of designs which are manual and automatic. The driver for manual transmission utilizes the clutch pedal and a gear stick that are either on the floor between the front seats or on the center console. Meanwhile, the electronic control unit or ECU in the automatic gearbox changes the gear ratio by using sensors to track the speed and engine revolutions per minute [7]. The clutch and manual transmission's typical functions are carried out automatically, permitting controlled slippage between the car's engine and transmission [8]. In urban driving conditions, automatic transmissions are generally more popular due to their convenience, which could influence their resale value compared to manual transmissions [9].

Age of car

Every car is subject to depreciation, but it has the greatest impact during the initial years of ownership, with the first year experiencing the steepest decline in value [10]. Depreciation is defined as the difference between the price of the car at the time of purchase and its current market value. For purchasers, awareness of depreciation can aid in accurately estimating long-term ownership costs and simplify negotiation processes [10]. The age of car is a significant factor influencing resale value [4]. As cars age, they may be perceived as less desirable due to absence of the latest features and safety technologies, resulting in lower resale values compared to newer models equipped with updated advancements.

Mileage

Mileage is defined as the total distance that a car has traveled since it was manufactured. Typically, it is measured in kilometres or miles. The maintenance requirements are significantly impacted by the car mileage [11]. In general, higher mileage made car parts more prone to damage and the frequently used parts like the engine, suspension, and braking system are more likely to experience wear and tear and lead to damage as the mileage increases [11]. The relationship between mileage and resale value is well-established, with lower mileage often associated with higher resale values [12]. A study also agreed that lower mileage will cost higher than a car with a high mileage, specifically on automatic transmission [3]. Hence, as mileage increases, cars need more frequent maintenance which will lead to lower resale value.

Engine capacity

A study indicated that engine capacity is a significant factor that affects a car's price [13]. A well-maintained car is one with a larger engine capacity, as it has the capability for higher speeds and better acceleration. Due to their superior performance characteristics, vehicles with larger engine capacity command higher resale values, which appeal to individuals who are seeking a more exhilarating driving experience. The resale value of the car will increase when the engine capacity is increased [14].

Body types

The body type of a car refers to the structure and design of a car’s exterior shell. It is an essential actor since it reflects a variety of aspects like functionality, market demand, and style. The common and popular body types in Malaysia are sedans because of their spaciousness, comfort, and affordability. An example of a Malaysian car with a sedan body type is Proton S70. In addition, SUV is also one of the body types such as Perodua Ativa [15]. It gained popularity among Malaysians due to its spacious interiors, higher ground clearance, and perceived safety features. Other body types included in Malaysia are Multi-purpose, MPV and hatchback. The most expensive body types of cars are sedans and SUVs [3].

Methodology

The population of the study was all Perodua-used cars listed on the Carro website and a sample consist of 156 used cars advertised on the Carro website from July 2024 until October 2024 was taken. The secondary data was obtained from https://carro.co/my/en using the self-key in data technique. The variables used for this study are the transmission types, the age of the car, the mileage, the car’s engine capacity, and the body types as shown in Table 1.

Table 1: Description of Variables

Variables	Scale of Measurement	Data Value
Resale Value (price in RM)	Ratio	
Transmission types	Nominal	0 = Automatic 1 = Manual
Mileage (km)	Ratio	
Engine Capacity	Ordinal	0 =>1300cc 1=1000cc – 1300cc
Age of car (month)	Ratio	
Body types	Nominal	0 = Hatchback 1 = Sedan 2 = SUV 3 = MPV

Constructing Dummy for Categorical Variables

Dummy variables are independent variables which take the value of either 0 or 1. In quantitative analysis, such as Multiple Linear Regression, a dummy variable is a numeric stand-in for qualitative variables or a logical proposition. Practically, it is simplest to understand the regression results when the dummy variable is limited to 0 and 1 [16].

For example, in this study, variable body types have four levels which were Sedan, SUV, MPV, and Hatchback. Its level of body type was recorded into three dummy variables with one type treated as the reference category. The reference category was hatchback since it had the highest category level in the dataset. The references for each categorical variable are shown in Table 2.

Table 2: Description of Dummy Variables

Variables	Reference Category	Value	Names of dummy
Transmission types	Automatic	0 = Otherwise 1 = Manual	Transmission group
Engine capacity	>1300cc	0 = Otherwise 1 = 1000cc-1300cc	Engine group
Body types	Hatchback	0 = Otherwise 1 = Sedan	Sedan
Body types	Hatchback	0 = Otherwise 1 = SUV	SUV
Body types	Hatchback	0 = Otherwise 1 = MPV	MPV

Data Analysis

Multiple linear regression was used to determine the significant factors that influenced the resale value of cars. In this subsection, procedures of Multiple Linear Regression (MLR) were discussed. The purpose is to determine if two or more independent variables (X1, X2, ..., X7) have a significant effect on the dependent variable (Y) [4]. The Multiple linear regression equation is given in equation 1.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_7 X_7 + \varepsilon \quad (1)$$

Where;

Y = resale value (RM)

β_0, β_1, β_2, β_3, β_4, β_5, β_6, and β_7 are the regression parameters or regression coefficients.

X_1 : Transmission types

X_2 : Mileage (km)

X_3 : Engine capacity

X_4 : Age of car (month)

X_5 : Body types (Sedan)

X_6 : Body types (SUV)

X_7 : Body types (MPV)

ε : random error

If the p-value of an independent variable is less than the significance level of 0.05, the variable is considered significant to the model, indicating that the factors are affecting the resale value of cars in Malaysia. During the model development process, the data were thoroughly examined, and the assumptions underlying the regression were validated, such as linearity between dependent and independent variables, normality of error terms, homoscedasticity, multicollinearity and autocorrelation, and outliers.

Results and Discussion

This research was conducted to identify the most significant factors affecting the resale value of cars in Malaysia. Multiple Linear Regression (MLR) was employed to fulfill the objective. All the assumptions of Multiple Linear Regression were met. ANOVA (Analysis of Variance) is used to assess the influence of one or more independent variables on a continuous dependent variable. The hypothesis test determines the statistical significance of the model. In Table 3, the p-value is 0.000, which is less than the significance level of 0.05, indicating that the model is statistically significant.

The coefficient of determination R^2 quantifies the proportion of variability in the dependent variable (Y) that is explained by the model. According to the value of R^2 presented in Table 3, 97.2% of the total variation in resale value (price in RM) can be explained by the independent variables, including transmission types, mileage, engine capacity, car age (in months), and body

types. The remaining 2.8% is attributed to other factors not included in the model. The model is a good fit for the data, as it captures most of the patterns and relationships in the dataset.

Table 3: Results of ANOVA and R-Squared

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	15440000000	7	2206417646	734.346	0.000
Residual	444681283.8	148	3004603.269		
Total	15890000000	155			
R-Squared (R^2)					0.972

Estimated Model

The model for multiple linear regression is as follows:

$$Y = 59468.351 - 12353.136X_1 - 0.025X_2 - 11315.296X_3 - 177.148X_4 + 6169.695X_5 + 22224.934X_6 + 9921.884X_7 \quad (2)$$

Where;

X_1 : Transmission types
X_2 : Mileage (km)
X_3 : Engine capacity
X_4 : Age of car (month)
X_5 : Body types (Sedan)
X_6 : Body types (SUV)
X_7 : Body types (MPV)

Testing of Model Parameters

After confirming the model's validity, the significance of each parameter was assessed. If the p-value of an independent variable is less than the significance level of 0.05, the variable is considered significant to the model. Table 4 shows that all independent variables are significant, as their p-values are all less than 0.05.

Table 4: Coefficient results

Model	Unstandardized coefficient	t	Sig.	Null Hypotheses
(Constant)	59468.351	110.387	0.000	
Transmission types	-12353.136	-14.999	0.000	H_0 rejected
Mileage (km)	-0.025	-3.395	0.001	H_0 rejected

Engine capacity	-11315.296	-16.402	0.000	H_0 rejected
Age of car (month)	-177.148	-18.559	0.000	H_0 rejected
Body types (Sedan)	6,169.695	8.376	0.000	H_0 rejected
Body types (SUV)	22224.934	26.827	0.000	H_0 rejected
Body types (MPV)	9921.884	28.860	0.000	H_0 rejected

The significant relationship between transmission types and resale value aligns with previous research by [17], which indicated that the price of manual cars is lower than automatic cars. Moreover, there is significant relationship between engine capacity and its resale value, highlighted that the cars with a bigger engines will cost more than the small engine capacity [18]. From this study, age of car (month) concluded to be the significant factor affecting cars resale value in Malaysia, supported by a study by [4]. In addition, there is significant relationship between mileage and resale value , which highlight that lower mileage typically corresponds to higher resale value [12], aligns with this study that showed mileage is one of the significant factors of car resale value. Furthermore, a car's resale value influenced by its body type is consistent with [19] study, where larger vehicles, such as MPVs and SUVs, generally have higher resale values compared to smaller cars like sedans and hatchbacks.

Acknowledgments

The authors would like to thank Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban for the facilities involved in making this research a success.

References

[1] Muti, S., Yıldız, K. (2023). Using linear regression for used car price prediction. *International Journal of Computational and Experimental Science and Engineering,* 9(1), 11–16.

[2] Abdullah, F., Rahman, M. A., Shidujaman, M., Hasan, M., Habib, M. T. (2023). Machine learning modeling for reconditioned car selling price prediction. *Seventh International Conference on Mechatronics and Intelligent Robotics*, 12779, 605–612.

[3] Barlybayev, A., Sankibayev, A., Kadyr, Y., Amangeldy, N., Sabyrov, T. (2023). Predicting used-vehicle resale value in developing markets: Application of machine learning models to the Kazakhstan car market. *Ingénierie des Systèmes d'Information*, 28(5), 1237–1246.

[4] Puteri, C., Safitri, L. (2020). Analysis of linear regression on used car sales in indonesia. *Journal of Physics: Conference Series,* 1469(1), 012143.

[5] Yesayan, L. A., Asryan, A. (2024). Predicting car prices based on secondary market data in Armenia (Undergraduate project report). Faculty of Business and Management, American University of Armenia.

[6] Institute, U. T. (2020). What is a transmission and how does it work? [December 7]. https://www.uti.edu/blog/automotive/transmission

[7] Palla, S. (2022). Types of car transmissions – different types of transmission in cars [November 30]. https://www.spinny.com/blog/types-of-car-transmissions/

[8] Karimjon, I., et al. (2022). Research of operational reliability of automatic trans- missions. *Texas Journal of Multidisciplinary Studies,* 8, 83–87.

[9] Ameen, H. A., Mahamad, A., Saon, S., Nor, D. M., Ghazi, K. (2020). A review on vehicle to vehicle communication system applications. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(1), 188–198.

[10] Motorway. (2024). Does age impact the value of your car. https://motorway.co.uk/sell-my-car/guides/does-age-impact-the-value-of-your-car

[11] Chen, F., Shang, D., Zhou, G., Ye, K., Ren, F. (2024). Mileage-aware for vehicle maintenance demand prediction. *Applied Sciences,* 14(16), 7341.

[12] Sun, Z. (2024). Research on factors affecting second-hand car market prices. *Theoretical and Natural Science*, 36, 128–135.

[13] Mammadov, H. (2021). Car price prediction in the USA by using liner regression. *International Journal of Economic Behavior (IJEB)*, 11(1), 99–108.

[14] Gollapalli, M., Alqahtani, T. A., Alhamed, D. H., Alnassar, M. R., Alajmi, A. M., Alali, Y. H., Abdulqader, M. M., Saadeldeen, A. (2023). Intelligent modelling techniques for predicting used cars prices in Saudi Arabia. *Mathematical Modelling of Engineering Problems*, 10(1),139-148

[15] Perodua. (2024). http://www.perodua.com.my/

[16] Achmad Rinaldo Fernandes, A., Solimun, & Nurjannah. (2022). Computational Statistics with Dummy Variables. *IntechOpen*. doi: 10.5772/intechopen.101460

[17] NationWide. (2021). Manual vs. automatic transmission: Which is better? [October 12]. https://blog.nationwide.com/vehicle/vehicle-buying-selling/manual vs-automatic-transmission

[18] Staff, C. (2022, September). What car engine size do i need?. https://www.carwow.co.uk/guides/choosing/what-engine- size-do-i-need-0521gref.

[19] Hafiz, H. (2022). Car depreciation rate malaysia: 8 important factors to consider [April 11]. https://www.qoala.my/en/blog/asset-management/car-depreciation- rate-malaysia/

Thermal Instability Analysis of Convection in a Horizontal Nanofluid Layer with Free-Free Boundary Condition

Norazuwin Najihah Mat Tahir[1,*], Nurfarissha Syazwina Muhamad Hairol[1], Noralia Qisti Mohammad Nazri[1]

[1]*Center of Mathematical Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARACawangan Negeri Sembilan, Kampus Seremban, Persiaran Seremban Tiga/1, 70300 Seremban, Negeri Sembilan*

[*]najihahtahir@uitm.edu.my

Abstract

Convective flows are commonly observed in geophysical systems and hold both fundamental and practical significance in engineering and technological applications. The initiation of convective flow whether stationary or oscillatory is known to cause defects such as striations, dendrites and bubbles in manufactured products. This study investigates the influence of viscous elasticity and an applied magnetic field on convective instabilities in a deep horizontal layer of viscoelastic nanofluid with free-free boundary conditions. A linear stability analysis was employed to identify the onset of both stationary and oscillatory instabilities. A closed-form expression for the critical Rayleigh number was derived using the Galerkin-typed weighted residuals method. The impact of the scaled stress relaxation parameter, scaled strain retardation parameter and Chandrasekhar number on system stability was analyzed. The results indicate that the magnetic field effectively delays the onset of both stationary and oscillatory convective instabilities in viscoelastic nanofluids.

Keywords*:* Stress Relaxation, Strain Retardation, Chandrasekhar Number, Free-Free Boundary Condition

Introduction

Nanofluids have attracted significant attention in recent years due to the growing need for high-performance thermal management systems. As a result, numerous studies have been carried out to better understand their thermophysical behavior [1]. The onset of convective instability in nanofluids has been explored in the works of Nield and Kuznetsov [2,3] and Yadav et al. [4]. The concept of nanofluids was first introduced by Choi [5], who defined them as engineered suspensions of nanoparticles (typically ranging from 1 to 100 nanometers) within a base fluid, such as water, ethylene glycol, or oil [6]. Ganguly et al. [7] similarly emphasized that nanofluids consist of uniformly distributed nanoparticles in various carrier liquids, including water and organic solvents.

The nanoparticles used in these fluids whether metallic or non-metallic such as Al_2O_3, CuO, Cu, SiO_2, and TiO_2, considerably enhance the heat transfer performance of the base fluid. This enhancement is primarily due to improved thermal conductivity, as discussed by Kakaç and Pramuanjaroenkij [8], making nanofluids highly effective in advanced thermal applications. A novel category of fluids known as nanofluids has emerged, offering significant potential for

enhancing thermophysical characteristics and improving heat transfer performance. Despite this, most investigations into non-Newtonian nanofluids have been experimental in nature [9], while theoretical analyses concerning flow instability have largely treated nanofluids as Newtonian fluids [10]. Experimental evidence, however, suggests that the Newtonian framework may not sufficiently capture the behavior of certain nanofluids [11]. Specifically, nanofluids with low nanoparticle concentrations tend to behave in a Newtonian manner, whereas higher concentrations lead to shear-thinning behavior, characteristic of non-Newtonian fluids.

The onset of convective instability in nanofluids was examined through linear stability analysis. Prior investigations by Yadav et al. [4], Nield and Kuznetsov [2,3] have explored this phenomenon in detail. Dastvareh and Azaiez [12] reported that the presence of nanoparticles and applied magnetic fields can significantly influence the stability characteristics of an initially unstable flow, either by amplifying or suppressing the onset of convection. Tahir and Kechil [13] conducted a linear stability analysis to examine the elasticity effects on hydromagnetic convective instability in a viscoelastic nanofluid layer as well as the magneto-convective instability in a horizontal viscoelastic nanofluid saturated porous layer by Tahir et al., [14]. Their findings also showed that magnetic fields contribute to the stabilization of the nanofluid layer in both stationary and oscillatory convection regimes. Additionally, it was observed that the intrinsic rheological properties of the fluid play a critical role in shaping the convective behavior.

Previous research has explored the thermal stability of nanofluid convection under various boundary conditions, including rigid–rigid and rigid–free configurations. However, the free–free boundary condition, which closely resembles conditions encountered in practical applications, has received comparatively limited attention [15]. In this configuration, both the upper and lower surfaces of the nanofluid layer are free to deform, resulting in distinct stability behavior. This added freedom introduces a broader spectrum of system responses and potential flow structures, which significantly influence heat transfer characteristics. Such behavior is particularly relevant for engineering applications involving solar thermal collectors, electronic cooling systems, and heat exchangers

Considering these observations, the present study focuses on exploring the non-Newtonian characteristics of nanofluids. It investigates how factors such as nanoparticle volume fraction, fluid viscosity and various external forces for both surface and body affect convective instability. Through theoretical analysis, this work aims to deepen the understanding of how these parameters influence the onset of stationary and oscillatory instabilities and to identify mechanisms that may either suppress or enhance such convective behaviors.

Problem Formulation

Considers a horizontal layer of viscoelastic non-Newtonian nanofluid subjected to a magnetic field, under free-free boundary conditions as shown in Figure 1. Models for nanofluids with negligible external forces were proposed by Buongiorno [16], while Nield and Kuznetsov [17] introduced a model for the natural convection onset in constant viscosity nanofluids. This study extends the framework of Narayana et al. [18] by considering the fluid as a non-Newtonian nanofluid.

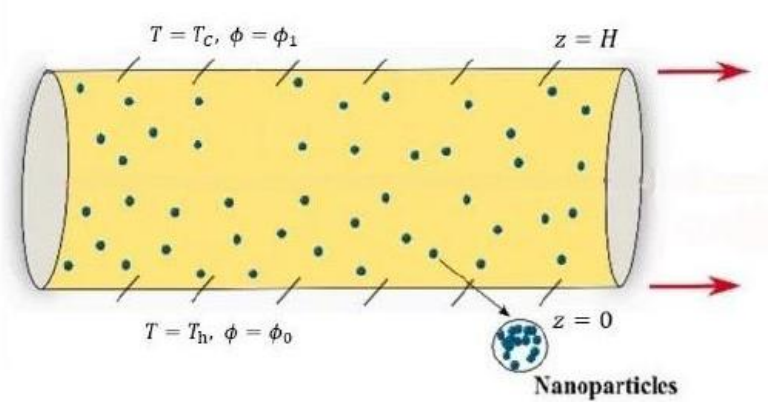


Figure 1: Schematic diagram of the physical system.

The formulation of the problem is based on the models proposed by Buongiorno [16], Narayana et al. [18] and Nield and Kuznetsov [17]. The basic governing equations for the mass, momentum, energy and nanoparticle conservation for the non-Newtonian nanofluid are

$$\nabla \cdot \mathbf{v} = 0, \tag{1}$$

$$\left(1 + \lambda_1 \frac{\partial}{\partial t}\right)\left[\rho_{f_0}\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v}\right) + \nabla p - \rho \mathbf{g} + \sigma u \mu_m^2 H_0^2 \boldsymbol{i} + \sigma v \mu_m^2 H_0^2 \boldsymbol{j}\right] = \mu\left(1 + \lambda_2 \frac{\partial}{\partial t}\right)\nabla^2 \mathbf{v}, \tag{2}$$

$$(\rho c)_f \left(\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T\right) = k\nabla^2 T + (\rho c)_p \left[D_B \nabla \phi \cdot \nabla T + \left(\frac{D_T}{T_c}\right)\nabla T \cdot \nabla T\right], \tag{3}$$

$$\frac{\partial \phi}{\partial t} + \boldsymbol{v} \cdot \nabla \phi = D_B \nabla^2 \phi + \left(\frac{D_T}{T_c}\right)\nabla^2 T, \tag{4}$$

where ∇ is divergence of a vector field, $\mathbf{v} = (u, v, w)$ is nanofluid velocity, λ_1 is relaxation time, λ_2 is retardation time, t is corresponds to time, ρ_{f_0} is nanofluid density at reference temperature, T_c, p is pressure, $\mathbf{g}$ is gravity, σ is electrical conductivity, μ_m is magnetic permeability, μ is coefficient viscosity, $(\rho c)_f$ is effective heat capacity of the fluid, k is thermal conductivity, $(\rho c)_p$ is effective heat capacity of the nanoparticle, D_B is Brownian diffusion coefficient, ϕ is nanoparticle volume fraction, T is temperature, D_T is thermophoretic diffusion coefficient and T_c is reference temperature. The nanofluid density, ρ is

$$\rho \cong \phi\rho_p + (1 - \phi)[\rho_{f_0} - \rho_{f_0}\beta(T - T_c)], \tag{5}$$

where ρ_p is nanoparticle mass density and β is volumetric coefficient of thermal expansion.

The boundary conditions are free at both lower and upper boundaries, given by

$$w = 0, \frac{\partial w}{\partial z} + a_1 H \frac{\partial^2 w}{\partial z^2} = 0, T = T_h, \phi = \phi_0 \text{ at } z = 0, \tag{6}$$

$$w = 0, \frac{\partial w}{\partial z} - a_2 H \frac{\partial^2 w}{\partial z^2} = 0, T = T_c, \phi = \phi_1 \text{ at } z = H. \tag{7}$$

Linear Stability Analysis

Linear stability analysis is applied to the system (1) – (7) to investigate the initiation of both stationary and oscillatory instabilities. Non-dimensionalization is carried out by choosing appropriate scaling parameters based on the fundamental dimensions present in the original governing equations. Following Narayana et al. [19] and Nield and Kuznetsov [17], the scaling variables for length, time, velocity, pressure, nanoparticle volume fraction and temperature are

$$(x^*, y^*, z^*) = \frac{(x, y, z)}{H}, t^* = \frac{\alpha}{H^2} t, (u^*, v^*, w^*) = (u, v, w) \frac{H}{\alpha_f}, p^* = \frac{H^2}{\mu \alpha_f} p,$$
$$\phi^* = \frac{\phi - \phi_0}{\phi_1 - \phi_0}, T^* = \frac{T - T_c}{T_h - T_c}, \tag{8}$$

where $\alpha_f = \frac{k}{(\rho c)_f}$, is the thermal diffusivity of the fluid, and asterisks denote dimensionless quantities. The resulting equations and boundary conditions are in the form of dimensionless partial differential equations. The non-dimensional equations after dropping the asterisks are

$$\nabla \cdot \mathbf{v} = 0, \tag{9}$$

$$\left(1 + \Lambda_1 \frac{\partial}{\partial t}\right)\left[\frac{1}{P_r}\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v}\right) + \nabla p + Rm\boldsymbol{k} - RaT\boldsymbol{k} + Rn\phi\boldsymbol{k} + Q(u\boldsymbol{i} + v\boldsymbol{j})\right]$$
$$= \left(1 + \Lambda_2 \frac{\partial}{\partial t}\right)\nabla^2 \mathbf{v}, \tag{10}$$

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = \nabla^2 T + \frac{N_B}{Le} \nabla \phi \cdot \nabla T + \frac{N_A N_B}{Le} \nabla T \cdot \nabla T, \tag{11}$$

$$\frac{\partial \phi}{\partial t} + \mathbf{v} \cdot \nabla \phi = \frac{1}{Le} \nabla^2 \phi + \frac{N_A}{Le} \nabla^2 T, \tag{12}$$

subject to the dimensionless boundary conditions

$$w = 0,\ \frac{\partial w}{\partial z} + a_1 \frac{\partial^2 w}{\partial z^2} = 0,\ T = 1,\ \phi = 0 \ \text{at}\ z = 0, \tag{13}$$

$$w = 0,\ \frac{\partial w}{\partial z} - a_2 \frac{\partial^2 w}{\partial z^2} = 0,\ T = 0,\ \phi = 1 \ \text{at}\ z = 1. \tag{14}$$

The nondimensional parameters are

$$\Lambda_1 = \lambda_1 \frac{\alpha_f}{H^2}, \Lambda_2 = \lambda_2 \frac{\alpha_f}{H^2}, P_r = \frac{\mu}{\rho_{f_0}\alpha_f}, Le = \frac{\alpha_f}{D_B}, Q = \frac{\sigma \mu_m^2 H_0^2 H^2}{\mu},$$

$$N_A = \frac{D_T(T_h - T_c)}{D_B T_c(\phi_1 - \phi_0)}, N_B = \frac{(\rho c)_p}{(\rho c)_f}(\phi_1 - \phi_0), Ra = \frac{[\rho_{f_0}\beta(T_h - T_c)]gH^3}{\mu\alpha_f}, \tag{15}$$

$$Rn = \frac{[(\rho_p - \rho_{f_0})(\phi_1 - \phi_0)]gH^3}{\mu\alpha_f}, Rm = \frac{[\rho_p \phi_0 + \rho_{f_0}(1 - \phi_0)]gH^3}{\mu\alpha_f},$$

where Λ_1 is the scaled stress relaxation parameter, Λ_2 is the scaled strain retardation parameter, P_r is the Prandtl number, Le is the Lewis number, Q is the Chandrasekhar number, N_A is the modified diffusivity ratio and N_B is the modified particle density increment.

Perturbation and Linearization

According to linear stability theory, minor perturbations are applied to the nanofluid layer to analyze its behavior. If these disturbances diminish with time, the system returns to its equilibrium state. However, if they grow, the system diverges from its steady state. The resulting linearized equations are presented below

$$\mathbf{v} = \mathbf{v}', p = p_b + p', T = T_b + T', \phi = \phi_b + \phi', \tag{16}$$

where prime denotes the perturbation quantity and p_b, T_b and ϕ_b are the pressure, temperature and concentration at the equilibrium state. Substituting the above expressions into the nondimensionalized partial differential equations (9) – (14) and linearizing the system yields a set of linear partial differential equations,

$$\left(1 + \Lambda_1 \frac{\partial}{\partial t}\right)\left[\frac{1}{P_r}\frac{\partial}{\partial t}\nabla^2 w' - \nabla_{\mathrm{H}}^2 RaT' + \nabla_{\mathrm{H}}^2 Rn\phi' + Q\nabla_{\mathrm{z}}^2 w'\right] = \left(1 + \Lambda_2 \frac{\partial}{\partial t}\right)\nabla^4 w', \tag{17}$$

$$\frac{\partial T'}{\partial t} - w' = \nabla^2 T' + \frac{N_B}{Le}\left(\frac{\partial T'}{\partial z} - \frac{\partial \phi'}{\partial z}\right) - 2\frac{N_A N_B}{Le}\frac{\partial T'}{\partial z}, \tag{18}$$

$$\frac{\partial \phi'}{\partial t} + w' = \frac{1}{Le}\nabla^2 \phi' + \frac{N_A}{Le}\nabla^2 T', \tag{19}$$

subject to boundary conditions,

$$w' = 0,\ \frac{\partial w}{\partial z} + a_1 \frac{\partial^2 w}{\partial z^2} = 0,\ T' = 0,\ \phi' = 0 \ \text{at}\ z = 0, \tag{20}$$

$$w' = 0,\ \frac{\partial w}{\partial z} - a_2 \frac{\partial^2 w}{\partial z^2} = 0,\ T' = 0,\ \phi' = 0 \ \text{at}\ z = 1. \tag{21}$$

where ∇^2 is the three-dimensional Laplacian operator, ∇_H^2 is the two-dimensional Laplacian operator, ∇_z^2 is the one-dimensional Laplacian operator with respect to $z-$plane and ∇^4 is the three-dimensional biharmonic operator.

Normal Modes

To transform the system of linear partial differential equations (17) – (21) into ordinary differential equations, the normal mode method is employed. The normal mode analysis utilizes two-dimensional periodic disturbances in an infinitely extended fluid layer [2,3,18], where the general solution is represented as a superposition of normal modes given by

$$(w', T', \phi') = [W(z), \Theta(z), \Phi(z)] e^{-(st + i\alpha_x + i\alpha_y)}, \tag{22}$$

where $W(z), \Theta(z)$ and $\Phi(z)$ are the amplitudes of the velocity, temperature and nanoparticles volume fraction, respectively. $\alpha = (\alpha_x^2 + \alpha_y^2)^{\frac{1}{2}}$ is the total wave number and s is a dimensionless complex growth rate where $\mathrm{Re}(s)$ is the growth rate and $\mathrm{Im}(s)$ is the frequency. For the stability at the marginal state $\mathrm{Re}(s) = 0$ when $\mathrm{Im}(s) = 0$, the stationary occurs and the convection is oscillatory when $\mathrm{Im}(s) \neq 0$. The system (17) – (21) in the form of the normal mode (22) are

$$(1 + \Lambda_1 s)\left[\frac{s}{P_r}(D^2 - \alpha^2)W + QD^2W + Ra\alpha^2\Theta - Rn\alpha^2\Phi\right] = (1 + \Lambda_2 s)(D^2 - \alpha^2)^2 W \tag{23}$$

$$W + \left(D^2 + \frac{N_B}{Le}D - 2\frac{N_A N_B}{Le}D - \alpha^2 - s\right)\Theta - \frac{N_B}{Le}D\Phi = 0, \tag{24}$$

$$W - \frac{N_A}{Le}(D^2 - \alpha^2)\Theta - \left[\frac{1}{Le}(D^2 - \alpha^2) - s\right]\Phi = 0. \tag{25}$$

The boundary conditions at $z = 0$ and $z = 1$ are thereby transformed into,

$$W = 0,\ DW + a_1 D^2 W = 0,\ \Theta = 0,\ \Phi = 0 \ \text{ at}\ z = 0, \tag{26}$$

$$W = 0,\ DW - a_2 D^2 W = 0,\ \Theta = 0,\ \Phi = 0 \ \text{ at}\ z = 1, \tag{27}$$

where $D = \frac{d}{dz}$. When $\Lambda_1 = \Lambda_2 = Q = 0,$ the system (23) – (25) reduced to the problem of Nield and Kuznetsov [18]. In this study, stress-free upper and lower boundaries will consider at $z = 0$ and $z = 1$,

$$W = 0, D^2 W = 0, \Theta = 0, \Phi = 0. \tag{28}$$

A feasible approach for solving the differential equations is the Galerkin-type weighted residual method. This technique addresses the linearized governing equations by representing the solution as a series expansion of selected trial functions. The functions W, Θ and Φ are expressed in the following form,

$$W = \sum_{n=1}^{N} A_n W_n \, , \Theta = \sum_{n=1}^{N} B_n \Theta_n , \Phi = \sum_{n=1}^{N} C_n \Phi_n , \tag{29}$$

where A_n, B_n, and C_n are unknown coefficients and $n = 1,2,3,\ldots,N$. Equations (23) – (25) are multiplied by W, Θ and Φ, respectively of (29). Applying integration by parts over the interval $z \in [0,1]$ results in a system containing $N-1$ unknowns. The trial functions that satisfied the boundary conditions (28) for free lower and upper boundaries are

$$W_1 = \Theta_1 = \Phi_1 = sin\, \pi z \, . \tag{30}$$

Results and Discussions

The thermal Rayleigh number, Ra serves as a critical parameter indicating the bifurcation point at which the system transitions from steady convection to mildly chaotic behavior and eventually to fully developed turbulence. The threshold for the onset of stationary convection $(\omega = 0)$ is given by

$$Ra_{stat} = \frac{(\pi^2 + \alpha^2)}{\alpha^2}[\pi^4 + 2\pi^2\alpha^2 + \alpha^4 + \pi^2 Q] - (N_A + Le)Rn. \tag{31}$$

The stationary instability is independent of Λ_1 and Λ_2 as the time-dependent terms become negligible in this regime. In the Newtonian fluid, where $Q = 0$ Eq. (31) simplifies to the stationary convection result reported by Nield and Kuznetsov [2]. For oscillatory convection $(\omega \neq 0)$, a complex growth rate $s = i\omega$ is considered. Consequently, the Rayleigh number takes the form $Ra = Ra_r + iRa_i$ where Ra_r and Ra_i denote the real and imaginary components, respectively. Enforcing the condition $Ra_i = 0$ yields the corresponding frequency ω associated with oscillatory instability given by

$$Ra_{osc} = \frac{1}{\alpha^2}\left\{\frac{\pi^4 + 2\pi^2\alpha^2 + \alpha^4}{1 + \Lambda_1^2\omega^2}[(\pi^2 + \alpha^2)(1 + \Lambda_1\Lambda_2\omega^2) - \omega^2(\Lambda_1 - \Lambda_2)] - \frac{\omega^2}{P_r}(\pi^2 + a + \pi^2 Q(\pi^2 + \alpha^2)\right\} - \frac{(\pi^2 + \alpha^2)^2(N_A + Le) + \omega^2 Le^2}{(\pi^2 + \alpha^2)^2 + \omega^2 Le^2} Rn. \tag{32}$$

Ra_{osc} (32) reduces to the form obtained by Nield and Kuznetsov [2] if $\Lambda_1 = \Lambda_2 = Q = 0$. This section highlights the influence of stress relaxation, strain retardation and magnetic field on both stationary and oscillatory instabilities. The neutral stability curves for the thermal Rayleigh number are presented in the corresponding parameter space, (α, Ra).

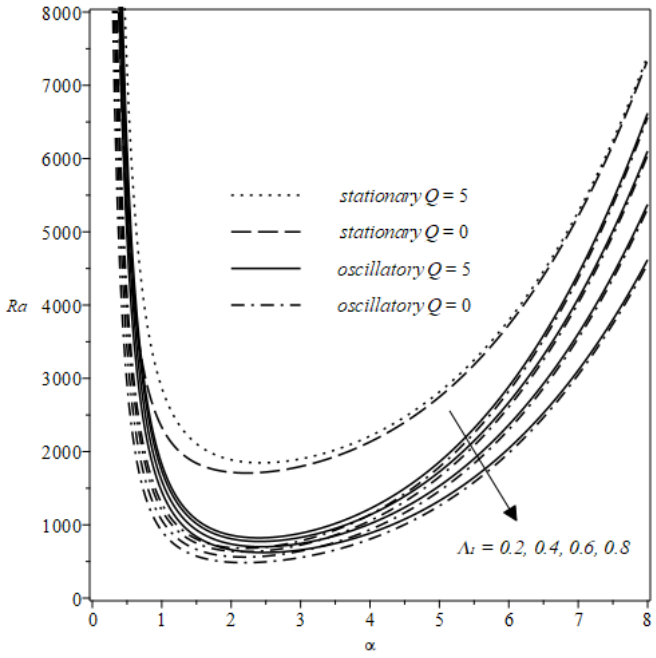


Figure 2: The neutral stability curves corresponding to stationary and oscillatory convection are presented for varying values of Λ_1, with fixed parameters $\Lambda_2 = 0.1, P_r = 5, Le = 100, N_A = 5, R_n = -10$ and $\omega = 1$.

Figure 2 illustrates the neutral stability curves representing the variation of the Rayleigh number, Ra with the wavenumber, α for different values of the stress relaxation parameter, Λ_1 under two magnetic field strengths, namely $Q = 0$ and $Q = 5$. The analysis includes both stationary and oscillatory modes of convective instability. The findings indicate that increasing Λ_1 leads to a reduction in the critical Rayleigh number for oscillatory convection, suggesting that higher values of Λ_1 promote an earlier onset of instability in the viscoelastic nanofluid layer. This behavior is attributed to the prolonged influence of applied stress on fluid elements, which diminishes the fluid's resistance to deformation. As Λ_1 increases, viscoelastic friction is reduced, lowering internal resistance and enabling convection to initiate at lower Rayleigh numbers.

Additionally, the critical Rayleigh number associated with stationary convection is observed to be higher than that of oscillatory convection. The enhanced stress relaxation effect intensifies the fluid's time-dependent response, making it more susceptible to oscillatory instability due to the reduced energy requirement for perturbation growth. In contrast, initiating stationary convection requires overcoming both viscous and elastic resistances, necessitating a higher Rayleigh number to trigger instability. Another key observation is the stabilizing influence of the magnetic field, the application of a magnetic field increases the critical Rayleigh number, indicating a suppression of convective onset.

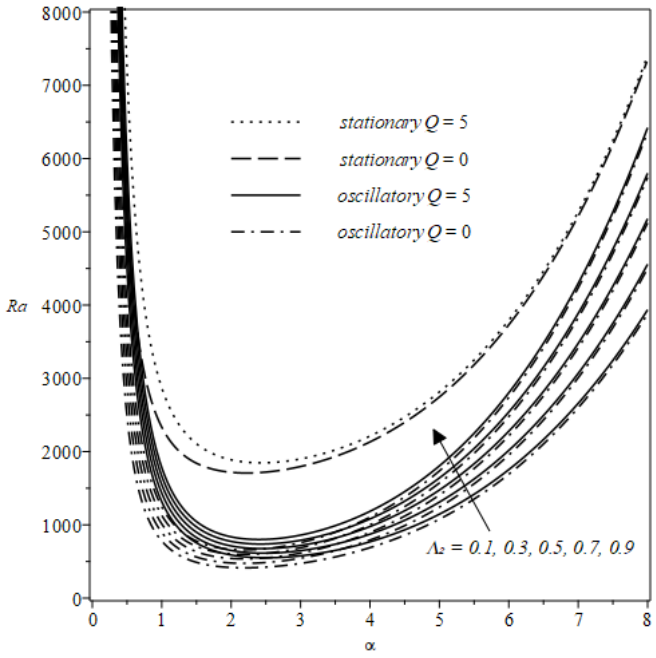


Figure 3: The neutral stability curves corresponding to stationary and oscillatory convection are presented for varying values of Λ_2, with fixed parameters $\Lambda_1 = 1.0, P_r = 5, Le = 100, N_A = 5, R_n = -10$ and $\omega = 1$.

Figure 3 displays the neutral stability curves illustrating the influence of the scaled strain retardation parameter, Λ_2 on the thermal Rayleigh number, Ra as a function of the wavenumber, α for various values of Λ_2. The results demonstrate that the critical Rayleigh number for oscillatory convection increases with an increase in Λ_2. This parameter, Λ_2 characterizes the dimensionless retardation time, representing the fluid's delayed stress response to deformation and accounting for its memory effect. A higher Λ_2 implies greater resistance to rapid changes in deformation, thereby exerting a stabilizing influence on the system. This stabilization delays the onset of convection, indicating that a greater thermal input is required to trigger instability in the viscoelastic nanofluid layer.

For both magnetic field strengths considered $Q = 0$ and $Q = 5$, the oscillatory mode consistently exhibits a lower critical Rayleigh number compared to the stationary mode. This suggests that oscillatory instabilities are more easily initiated, requiring less thermal energy, while the stationary modes demand a higher Rayleigh number, indicating comparatively greater stability in the fluid system.

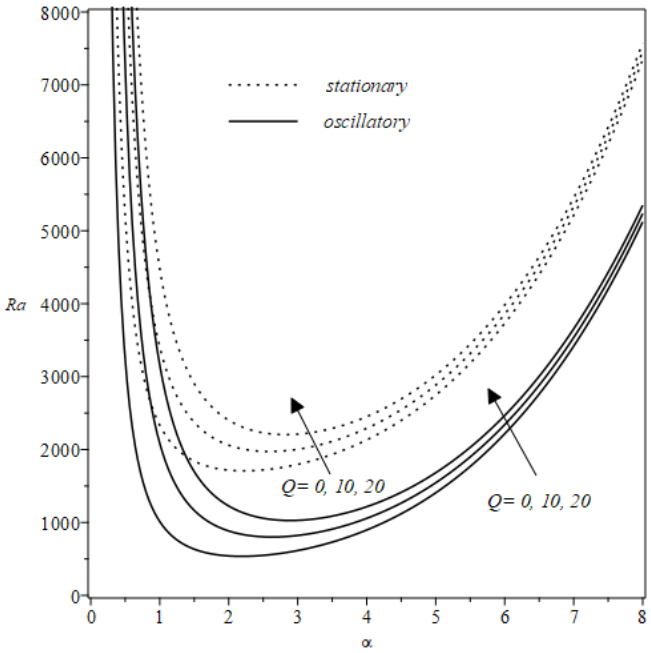


Figure 4: The neutral stability curves corresponding to stationary and oscillatory convection are presented for varying values of Q, with fixed parameters $\Lambda_1 = 1.0, \Lambda_2 = 0.1, P_r = 5, Le = 100, N_A = 5, R_n = -10$ and $\omega = 1$.

Figure 4 demonstrates that the thermal Rayleigh number increases with the Chandrasekhar number, Q for both stationary and oscillatory convection modes. As Q increases, corresponding increases in the Rayleigh number indicate enhanced thermal stability, implying that a stronger temperature gradient is required to initiate convective motion. This stabilizing effect is primarily attributed to the influence of the magnetic field, quantified by the Chandrasekhar number, which governs the strength of the Lorentz force acting within the fluid.

The Chandrasekhar number represents the ratio of magnetic to viscous forces, and a higher value signifies a more dominant magnetic influence. An intensified Lorentz force generates additional resistive drag, thereby suppressing fluid motion and increasing the threshold for the onset of convection. Consequently, the presence of a stronger magnetic field leads to reduced flow velocity and greater resistance to convective instability.

Conclusion

This study examines the convective instability in a horizontal layer of nanofluid bounded by free surfaces. By employing linear stability analysis, the governing nonlinear partial differential equations (PDEs) are transformed into a system of ordinary differential equations (ODEs) through normal mode analysis. The investigation focuses on the influence of key dimensionless parameters namely, the scaled stress relaxation parameter, the scaled strain retardation

parameter and Chandrasekhar number on both stationary and oscillatory convection modes. To verify the validity of the results, comparisons were made with existing literature, and the outcomes show strong agreement with previous findings, thereby confirming the accuracy of the present analysis. The main findings are summarized as follows,

- For fixed values of $\Lambda_2 = 0.1, P_r = 5, Le = 100, N_A = 5, R_n = -10$ and $\omega = 1$, an increase in the scaled stress relaxation parameter, Λ_1 leads to a decrease in the critical Rayleigh number for oscillatory convection, indicating earlier onset of instability.
- Under the same conditions, increasing the scaled strain retardation parameter, Λ_2 results in a higher oscillatory thermal Rayleigh number, suggesting enhanced system stability due to delayed convective onset.
- For both stationary and oscillatory convection modes, the thermal Rayleigh number increases with the Chandrasekhar number, Q given $\Lambda_1 = 1.0, \Lambda_2 = 0.1$ and the same values for the remaining parameters. This highlights the stabilizing effect of the magnetic field through the influence of the Lorentz force.

Acknowledgments

The authors would like to express their sincere gratitude to Universiti Teknologi MARA (UiTM) for the facilities, resources and continuous support provided throughout the completion of this research. Appreciation is also extended to colleagues and peers for their valuable insights and constructive discussions that contributed to the improvement of this work.

References

[1] Patil, M. S., Seo, J. H., Kang, S. J. and Lee, M. Y. (2016). Review on synthesis, thermo-physical property and heat transfer mechanism of nanofluids. *Energies*, 9(10), 840. https://doi.org/10.3390/en9100840

[2] Nield, D. A. and Kuznetsov, A. V. (2010). The onset of convection in a horizontal nanofluid layer of finite depth. *European Journal of Mechanics B/Fluids*, 29(3), 217–223. https://doi.org/10.1016/j.euromechflu.2010.02.003

[3] Nield, D. A. and Kuznetsov, A. V. (2012). The onset of convection in a layer of a porous medium saturated by a nanofluid: Effects of conductivity and viscosity variation and cross-diffusion. *Transport in Porous Media*, 92, 837-846. https://doi.org/10.1007/s11242-011-9935-2

[4] Yadav, D., Bhargava, R., Agrawal, G. S., Yadav, N., Lee, J. and Kim, M. C. (2014). Thermal instability in a rotating porous layer saturated by a non-Newtonian nanofluid with thermal conductivity and viscosity variation. *Microfluidics and Nanofluidics*, 16, 425-440. https://doi.org/10.1007/s10404-013-1234-5

[5] Choi, S. U. and Eastman, J. A. (1995). *Enhancing thermal conductivity of fluids with nanoparticles* (No. ANL/MSD/CP-84938; CONF-951135-29). Argonne National Lab.(ANL), Argonne, IL (United States).

[6] Buongiorno, J., Venerus, D. C., Prabhat, N., McKrell, T., Townsend, J., Christianson, R., Tolmachev, Y. V., Keblinski, P., Hu, L. W., Alvarado, J., Bang, I. C., Bishnoi, S. W., Bonetti, M., Botz, F., Cecere, A., Chang, Y., Chen, G., Chen, H., Chung, S. J., … Zhou, S. Q. (2009). A benchmark study on the thermal conductivity of nanofluids. *Journal of Applied Physics, 106*(9). https://doi.org/10.1063/1.3245330

[7] Ganguly, S., Sikdar, S. and Basu, S. (2009). Experimental investigation of the effective electrical conductivity of aluminum oxide nanofluids. *Powder Technology*, 196(3), 326-330. https://doi.org/10.1016/j.powtec.2009.08.010

[8] Kakaç, S. and Pramuanjaroenkij, A. (2009). Review of convective heat transfer enhancement with nanofluids. *International Journal of Heat and Mass Transfer,* 52(13-14), 3187-3196. https://doi.org/10.1016/j.ijheatmasstransfer.2009.02.006

[9] Ling, Z., He, Z., Xu, T., Fang, X., Gao, X. and Zhang, Z. (2017). Experimental and numerical investigation on non-Newtonian nanofluids flowing in shell side of helical baffled heat exchanger combined with elliptic tubes. *Applied Sciences*, 7(1), 48. https://doi.org/10.3390/app7010048

[10] Chand, R., Rana, G. and Yadav, D. (2017). Thermal instability in a layer of couple stress nanofluid saturated porous medium. *Journal of Theoretical and Applied Mechanics*, 47(1), 69–84. https://doi.org/10.1515/jtam-2017-0005

[11] Yadav, D., Mohamed, R. A., Lee, J. and Cho, H. H. (2017). Thermal convection in a Kuvshiniski viscoelastic nanofluid saturated porous layer. *Ain Shams Engineering Journal*, 8(4), 613–621. https://doi.org/10.1016/j.asej.2015.11.023

[12] Dastvareh, B. and Azaiez, J. (2017). Instabilities of nanofluid flow displacements in porous media. *Physics of Fluids*, 29(4). https://doi.org/10.1063/1.4978890

[13] Tahir, N. N. M. and Kechil, S. A. (2019). Elasticity effects on hydromagnetic convective instability in viscoelastic nanofluid layer. *ASM Science Journal*, 87 – 93.

[14] Tahir, N. N. M., Ishak, F. and Kechil, S. A. (2016). Magneto-convective instability in a horizontal viscoelastic nanofluid saturated porous layer. *Jurnal Teknologi (Sciences & Engineering)*, 78(12-3).

[15] Khalid, I. K., Mokhtar, N. F. M. and Ibrahim, Z. B. (2021). Rayleigh-Benard convection in nanofluids layer saturated in a rotating anisotropic porous medium with feedback control and internal heat source. *CFD Lett.*, 13(11), 1-20. https://doi.org/10.37934/cfdl.13.11.120

[16] Buongiorno, J. (2006). Convective transport in nanofluids. *ASME Journal Heat and Mass Transfer,* 128(3), 240 – 250. https://doi.org/10.1115/1.2150834

[17] Nield, D. A. and Kuznetsov, A. V. (2014). Thermal instability in a porous medium layer saturated by a nanofluid: a revised model. *International Journal of Heat and Mass Transfer*, *68*, 211-214. https://doi.org/10.1016/j.ijheatmasstransfer.2013.09.026

[18] Narayana, M., Gaikwad, S. N., Sibanda, P. and Malge, R. B. (2013). Double diffusive magneto-convection in viscoelastic fluids. *International Journal of Heat and Mass Transfer*, *67*, 194-201. https://doi.org/10.1016/j.ijheatmasstransfer.2013.08.027

The Relationship of Excessive Online Gaming on Academic and Lifestyle Outcomes Among University Students

Nur Alya Binti Ahmad Shafri[1], Nur Afiqah Aliya Binti Mohd Nizam[1], Anis Fathiah Binti Afandi[1], Hafizul Izham Bin Mohamad Rizal[1], Muhammad Syafiq Bin Jaafar[1], Suriyati Ujang[1,*]

[1]*Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Cawangan Pahang, Kampus Raub, 27600 Pahang, Malaysia*

[*]suriyatiujang@uitm.edu.my

Abstract

This study investigates how time spent on online gaming affects university students' academic performance, stress levels, life skills, social interaction, and sleep schedule. A total of 338 students from five academic programs at UiTM Pahang, Raub Campus, participated in the survey. Data were collected using structured questionnaires and analysed using descriptive statistics and Spearman's correlation analysis. Results show that students who play less than one hour daily tend to have better academic performance and life skills. However, longer gaming durations are weakly to moderately linked to increased stress, disrupted sleep, and slight social changes. These findings highlight the importance of balancing gaming habits and digital wellness education.

Keywords*:* Online gaming, academic performance, stress, life skills, sleep patterns

Introduction

Online games are video games played over the Internet or a computer network. These games span various genres, including first-person shooters, strategy games, and role-playing games, and are accessible on personal computers, gaming consoles, and mobile devices. Many young people, particularly university students, are drawn to online gaming for entertainment and as a way to relieve academic stress.

However, growing evidence shows that excessive time spent on online games can have negative effects on students [1]. As games evolve in graphics and gameplay, they become more engaging, making it easy for students to lose track of time. Being engrossed in gaming would result in reduced study hours, missed classes, and neglected academic responsibilities. For instance, one study [2] discovered that Australian schoolchildren who played electronic games for more than two hours a day tended to perform poorly on reading and numeracy assessments, indicating a connection between gaming and a lack of focus in the classroom. These consequences may affect students' general well-being in addition to their academic performance.

Academic performance is a key measure of student success and usually, usually assessed through exam results, coursework, and classroom participation. Students who dedicate excessive time to gaming often struggle academically due to reduced time for studying, might miss scheduled class time and even struggle to focus in class. However, when digital game-based learning incorporates educational content and games, it has demonstrated potential for enhancing cognitive skills including critical thinking, memory, and problem-solving [3-4].

Mental health is also a significant area of concern. Students who are addicted to online gaming may experience elevated stress levels, particularly when exposed to violent or competitive content. Shin and Ahn [5] discovered that violent video games had significantly amplified the emotional reactivity and aggressive thoughts in teenagers, implying that repeated exposure can exacerbate stress and impair their focus. Yoo [6] identified five distinct patterns of gaming behaviour among adolescents and linked them to psychosocial factors such as academic stress, poor self-control, loneliness, and lack of peer support. These findings suggest that unrestrained gaming is both a symptom and a cause of psychological stress. This emphasises the complex interplay between emotional health and digital behaviour in student populations.

When students prioritise games over relationships, their social interactions can suffer as well. A lack of family or peer engagement can result in emotional dependence on games and frayed social bonds. Uncontrolled gaming time would also impact their sleep, as students frequently stay up late gaming, resulting in poor sleep quality and afternoon fatigue. Studies have found a clear correlation between sleep deprivation and poor academic performance [7-8]. Although Internet games are intended for entertainment, many students, particularly in poorer nations, exhibit signs of gaming addiction. This habit can jeopardise their academic achievement, mental health, social relationships, time management, and sleep patterns. Given that students are among the most regular users of online games, it is critical to understand the implications of excessive gaming and to devise preventive measures.

Therefore, the current study aims to examine the impact of time spent on online gaming on various aspects of students' lives. Specifically, it seeks to investigate the relationship between time spent on online games and students' academic performance, stress levels, life skills, socialisation, and sleep patterns. The study focuses on students at UiTM Pahang, Raub Campus and aims to contribute valuable insights for educators and policymakers to better support students' well-being and academic success.

Methodology

This study involved a total population of 2,479 students from Universiti Teknologi MARA Pahang, Raub Campus. The campus offers five academic programs: AM110 (905 students), BA119 (500 students), BA111 (557 students), CS110 (312 students), and CS111 (205). A cross-sectional study design was chosen to investigate the effects of excessive time spent on online gaming among students at a specific point in time. Data were collected using a standardised questionnaire distributed to selected students from various programs.

Measuring Instrument

The main instrument used for data collection was a self-administered questionnaire, developed to explore the relationship between online gaming and various aspects of student life. The questionnaire was structured into six sections, labelled Section A to Section F, each aligned with the study's respective research objectives. The instrument was designed to collect both demographic information and data related to the behavioural, academic, and psychological effects of online gaming. Section A addresses background information, while Sections B to F measure specific variables using a five-point Likert scale ranging from *"strongly disagree"* to *"strongly agree"*. A summary of each section is provided in Table 1.

Table 1: Overview of questionnaire sections

Section	Content Description
A	Demographic information
B	Academic Performance
C	Stress Level
D	Life-Based Skills
E	Socialization
F	Sleep Schedule

Sampling Technique

A stratified sampling technique was employed to ensure that all academic programs at UiTM Pahang, Raub Campus were proportionally represented in the study. Stratified sampling involves dividing the population into subgroups or strata that share similar characteristics. In this case, students enrolled in five different academic programs acted as strata. For a population of 2,479 students, the total number of

students enrolled at UiTM Pahang Branch, Raub Campus, according to [9], the appropriate sample size is 331 respondents. The sample size for each course was calculated using a stratified sampling formula based on the proportion of students in each program relative to the total population. The sample for each Course is listed in Table 2.

Table 2: Number of respondents by course

Course Code	Program Name	Number of Respondents
AM110	Public Administration	121
BA119	Banking Studies	67
BA111	Business and Management	74
CS110	Computer Science	42
CS111	Statistics	27

This stratified approach is hoped to improve the sample's representativeness and enable more accurate comparisons of students from various academic disciplines related to gaming, academic performance and lifestyle.

Procedure for Data Analysis

The Cronbach's Alpha was used in this study to determine the reliability of the questionnaire. An alpha value of 0.70 or higher is often deemed satisfactory, suggesting that the instrument produces reliable results [10,11]. In this study, descriptive statistics were used to summarise and interpret the basic characteristics of the collected data. This included using statistical measures such as mean, standard deviation, frequency, and percentage, as well as visual representations such as tables and bar charts. The descriptive analysis provided a clear overview of the demographic characteristics and response distribution, allowing for the identification of patterns and trends in the amount of time students spent playing online games.

To investigate the relationships between variables, the study employed Spearman's Rank Correlation Coefficient, a non-parametric measure that assesses the strength and direction of the association between two ranked variables. This strategy was appropriate given the ordinal structure of the Likert-scale responses. In this study, the independent variable was time spent playing online games, while the dependent variables were academic performance, stress level, life-based skills, socialisation, and sleep schedule. Spearman's correlation was used to assess the research goals, evaluating whether there are significant associations between excessive gaming and each of these characteristics among students in the selected academic programs at UiTM Pahang, Raub Campus.

Results and Discussion

Descriptive Analysis

Cronbach’s alpha was used to determine the internal consistency of the instrument. According to Lavrakas [12], a Cronbach’s alpha value of 0.80 or above is generally considered acceptable, while values between 0.90 and 0.95 are deemed excellent. In this study, the overall Cronbach’s alpha was 0.927, indicating a high level of reliability for the instrument used. Table 3 summarises the reliability values for each variable in the study.

Table 3: Number of respondents by course

Variables	Cronbach's Alpha	Number of Items
Academic Performance	0.787	7
Stress Level	0.858	7
Life-Based Skills	0.919	7
Socialization	0.742	10
Sleep Schedule	0.943	8

The results demonstrate that all constructs have acceptable to excellent levels of internal consistency, with the highest reliability found in the sleep schedule variable (α = 0.943) and the lowest in socialisation (α = 0.742), which is still within acceptable limits.

Gender and Preference of Online Games

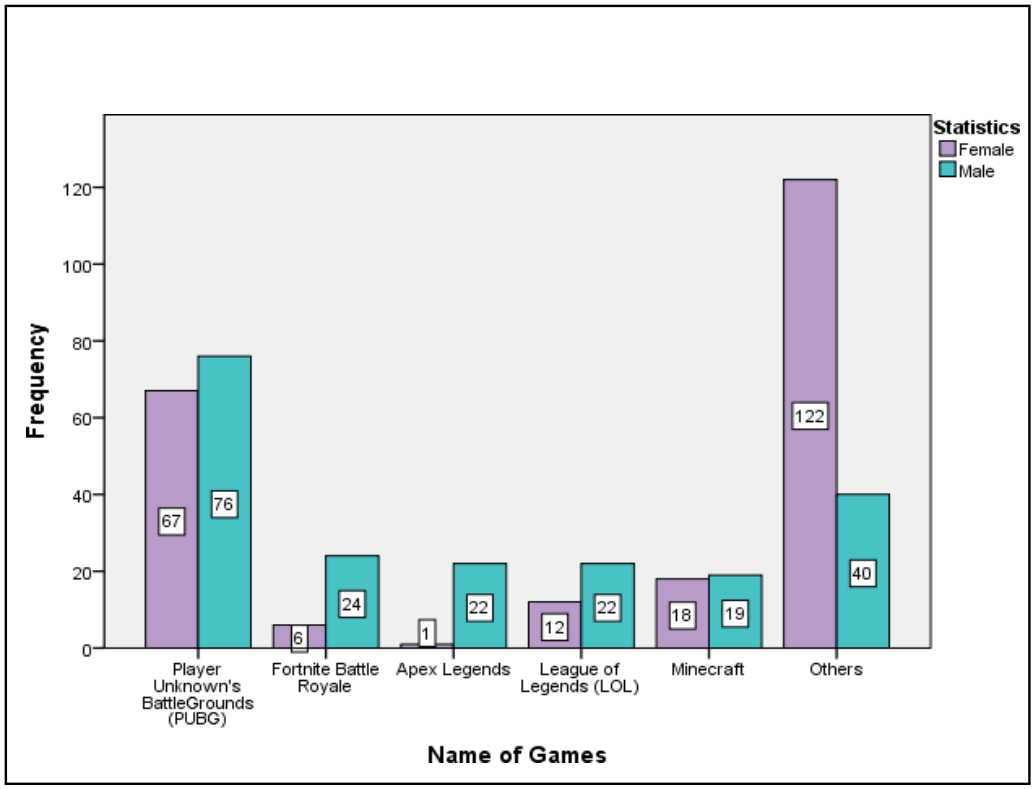


Figure 1: Frequency of respondents' preference for online games by gender

Figure 1 illustrates the types of online games played by students, categorised by gender. Among the games listed, Player Unknown's Battlegrounds (PUBG) was the most popular, with 76 male and 67 female respondents indicating it as their preferred game. This suggests high overall engagement with PUBG among both genders.

Male respondents showed a stronger preference for competitive, battle-based games such as Fortnite Battle Royale (n = 24), Apex Legends (n = 22), and League of Legends (LOL) (n = 22), compared to female respondents, whose frequencies for these games were relatively low: 6, 1, and 12, respectively. The minimal female representation in these categories suggests that these games are more favoured by male players.

In contrast, Minecraft showed a relatively balanced interest between genders, with 19 male and 18 female respondents indicating it as their game of choice. The most striking gender difference appeared in the "Others" category, where 122 female respondents compared to 40 males selected unspecified or

less common games, compared to 40 males. This suggests that female students have more diverse gaming preferences or engage in casual games not represented among the mainstream titles listed. The data indicates that male students tend to favour more competitive, action-based games, while female students demonstrate a more diverse range of game preferences, as shown by their dominant selection of the "Others" category.

Program Code and Frequency of Online Gaming

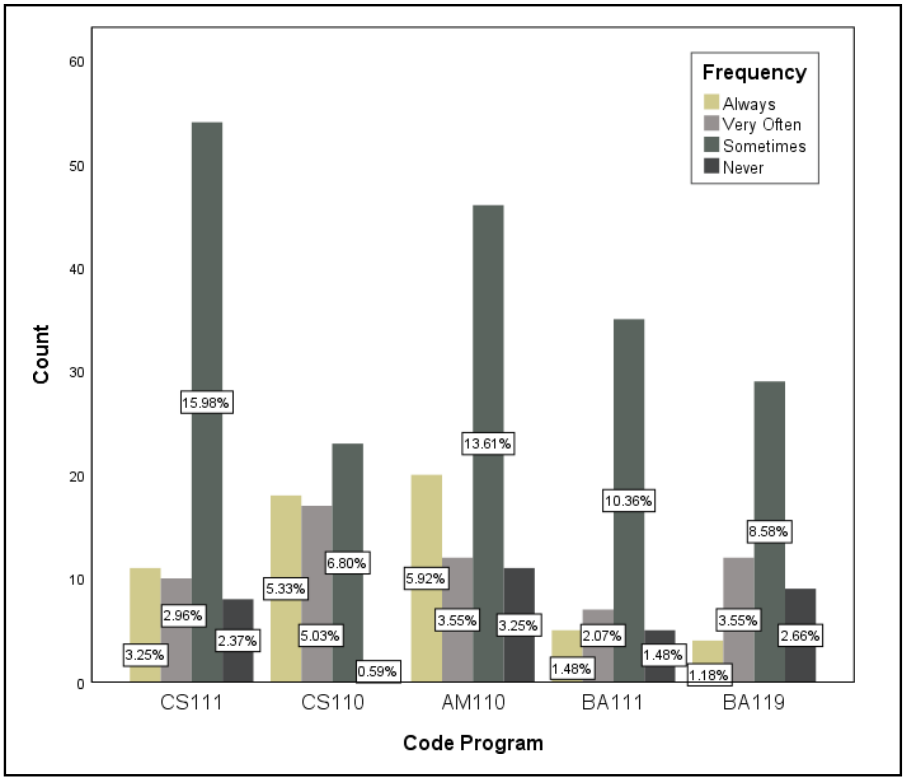


Figure 2: Frequency of respondents' course by frequency of online gaming

Figure 2 shows the frequency of online gaming frequency across different academic programs: CS111, CS110, AM110, BA111, and BA119, using four frequency categories: Always, Very Often, Sometimes, and Never.

The data reveals that "Sometimes" was the most common response across all programs. CS111 students constituted the highest proportion of occasional gamers, with 15.98% indicating they played online games "Sometimes." This is followed by AM110 students (13.61%) and BA111students (10.36%), suggesting a prevalent but moderate level of gaming activity in these programs.

In contrast, the "Never" category also shows a significant number of responses, particularly in CS110 (0.59%) and AM110 (3.25%), indicating that a notable group of students in these programs do not engage in online gaming at all. The "Always" and "Very Often" categories received comparatively fewer responses across all the programs. However, the CS110 and AM110 programs showed a higher tendency toward frequent gaming, with 5.33% and 5.92%, respectively, reporting that they "Always" played online games. Overall, the results suggest that while occasional gaming is common across all programs, particularly among students in CS111, AM110, and BA111, frequent gaming (Always/Very Often) remains limited to smaller groups, and a moderate number of students abstained from online gaming altogether.

Gender, Online Gaming, and Academic Performance

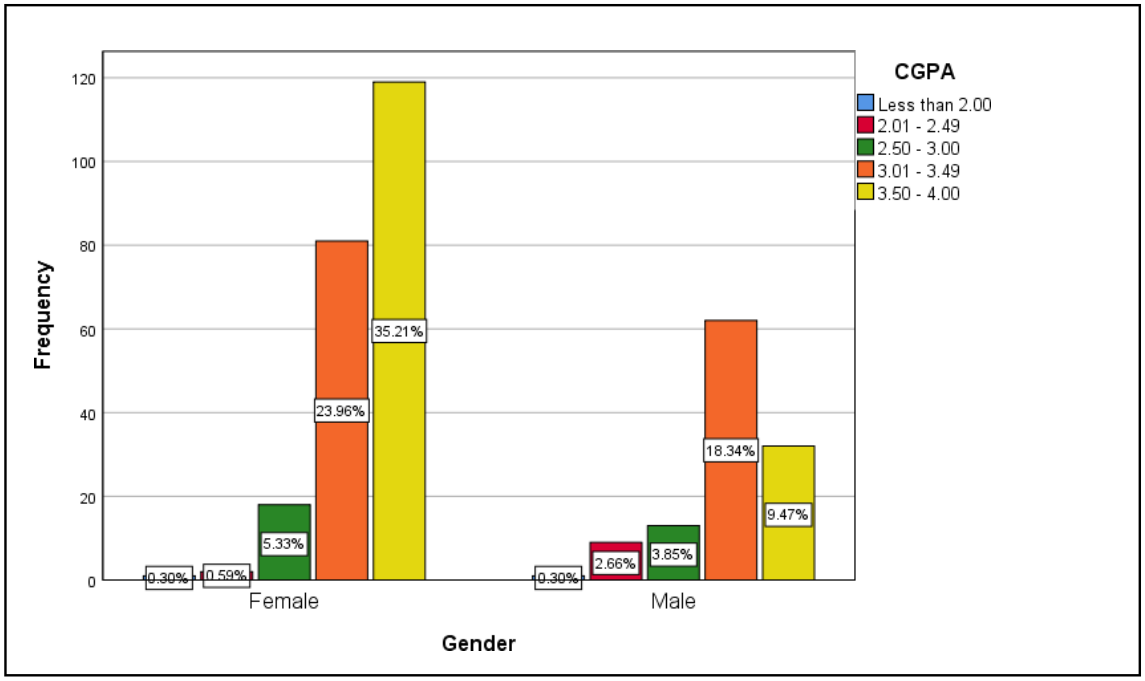


Figure 3: Frequency of respondents' CGPA by academic performance

Figure 3 illustrates the distribution of CGPA among male and female respondents who played online games. A significant proportion of respondents (44.68%) who achieved a CGPA between 3.50 and 4.00 were female (35.21%), whereas only 9.47% were male. Female students also dominated the other CGPA categories (3.01–3.49 and 2.50–3.00). Very few respondents, equally split by gender (0.30%), scored below 2.00. Overall, female students who played online games tended to perform better academically than their male counterparts.

Duration of Online Gaming and Academic Performance

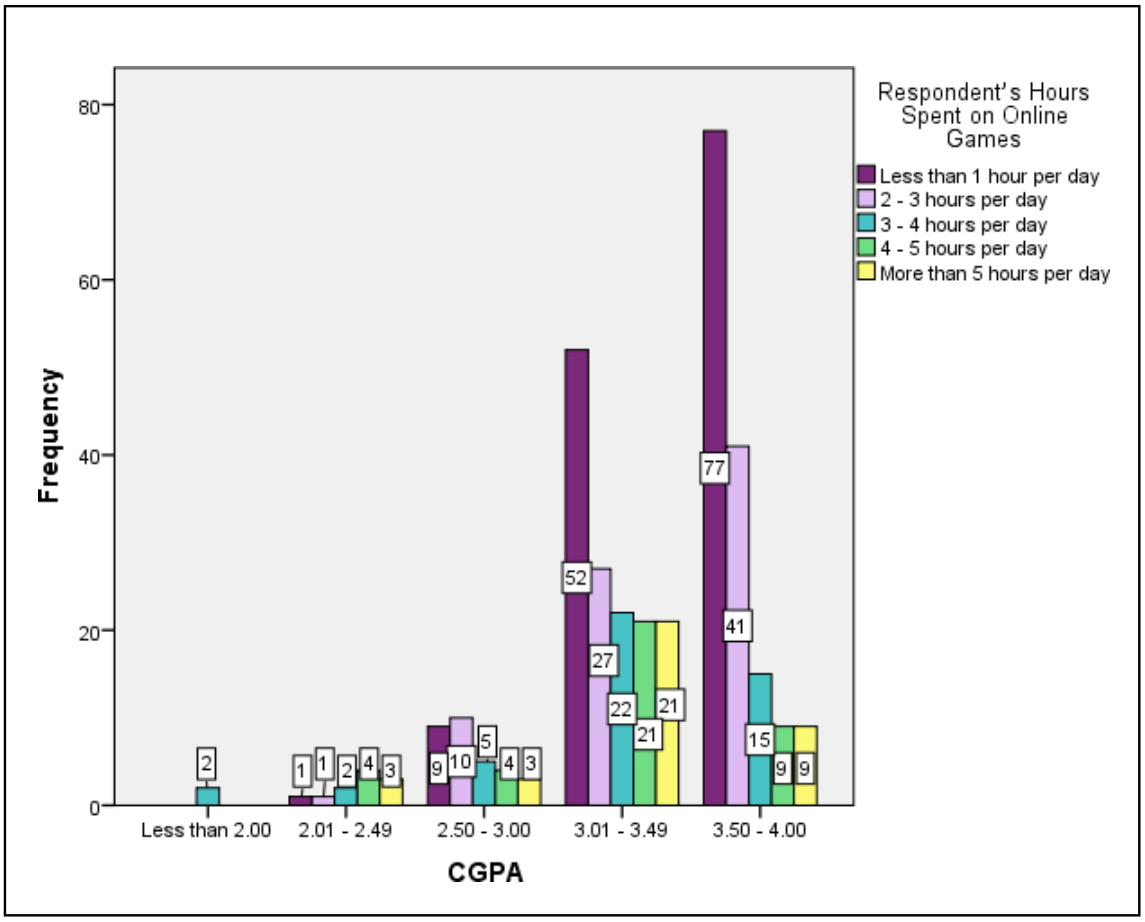


Figure 4: Frequency of respondents' CGPA by hours spent on online games

Figure 4 further illustrates the relationship between CGPA and gaming duration. A clear trend emerges where students who spend less than 1 hour per day on online games most frequently fall within the higher CGPA categories (3.01–3.49 and 3.50–4.00), with frequencies of 52 and 77, respectively. Conversely, students with lower CGPAs (below 3.00) are more evenly spread across all gaming durations, including those who play 3 or more hours per day. Notably, very few students with CGPAs below 2.50 reported gaming for less than one hour. This visual reinforces earlier statistical findings that limiting gaming time is associated with better academic performance.

Table 4 summarises the average scores of students' perceptions of the effects of online gaming across five measured dimensions.

Table 4: Summary of mean scores

Variables	Mean Score	Interpretation
Academic Performance	2.76	Neutral
Stress Level	2.94	Neutral
Life-Based Skills	3.19	Slightly Positive
Socialization	2.97	Neutral to Positive
Sleep Schedule	2.91	Neutral

In terms of self-reported perceptions, the students generally held neutral views regarding the impact of online gaming across most domains. They agreed that their grades remained satisfactory despite gaming and acknowledged some skill-building aspects such as multitasking and problem- solving. Although the students maintained positive relationships with peers and family, minor sleep disturbances and stress related to gaming were reported. Overall, the scores suggest that the perceived impacts of online gaming are mostly neutral, with some areas showing mild benefits or manageable drawbacks.

The Spearman's correlation analysis indicated a strong positive relationship between limited gaming time and higher academic performance. Longer gaming hours showed weak to moderate correlations with elevated stress, diminished sleep quality, and slight social impacts.

Inferential Analysis

Spearman's Correlation Coefficient was used to evaluate the strength and direction of the relationships between time spent on online games and the five dependent variables.

Table 5: Summary of correlation coefficient results

Variables	Correlation Coefficient (r)	Strength	Significance ($p < 0.01$)
Academic Performance	0.543	Strong Positive	Yes
Stress Level	0.346	Weak Positive	Yes
Life-Based Skills	0.347	Weak Positive	Yes
Socialization	0.417	Moderate Positive	Yes
Sleep Schedule	0.473	Moderate Positive	Yes

All variables reported in Table 5 demonstrated statistically significant positive correlations with gaming duration. The strongest correlation was observed between academic performance and gaming time ($r = 0.543$), whereas weaker relationships were observed between stress and life-based skills. Moderate correlations were identified between socialisation and sleep schedule.

These findings align with those of previous studies suggesting that moderate online gaming does not necessarily harm students' academic or social development. Students who limit their gaming time to less than an hour per day demonstrate academic and skill-based improvement, supporting claims for digital learning and cognitive engagement. However, the modest to moderate beneficial relationships shown between prolonged gaming sessions and stress, interrupted sleep, and social pressure highlight the need for time management. These effects, albeit not severe, underline the importance of raising awareness of about the potential negative consequences of unlicensed gaming.

Overall, this study supports a balanced approach to gaming, where students can benefit from its cognitive and recreational aspects without compromising their academic responsibilities and well-being with profound self-control. Universities and educators should also consider integrating digital wellness education and time management training into student development programs to instil responsible gaming behaviours.

Conclusion

This study investigated five key areas, academic performance, stress level, life- based skills, socialisation, and sleep schedule, and whether they are affected by excessive time spent on online gaming among university students. The findings reveal that while online gaming is a common activity for entertainment among students, the impact varies depending on the intensity and duration of gameplay across certain academic programs.

The results from both descriptive and inferential analyses suggest that moderate exercise engagement with adequate time management in online gaming does not necessarily hinder academic success or social well-being. A significant number of students who played for less than one hour per day achieved high academic performance. Furthermore, respondents recognised the potential benefits of gaming,

particularly in improving problem-solving and multitasking abilities.

However, the study also identified weak to moderate positive correlations between excessive gaming and elevated stress levels, disrupted sleep schedules, and minor changes in social behaviour. These findings highlight the importance of maintaining a balanced approach to online gaming to mitigate the potential negative effects of excessive gaming

In conclusion, online gaming, when practised in moderation, may serve as a constructive activity that complements students' development. Educational institutions and stakeholders should consider incorporating digital literacy and time management strategies into student development programs to foster healthier gaming habits. With better time management, it is hoped that students will have a better quality of life without excluding the fun they can have. Future research should expand on upon these findings using longitudinal designs or qualitative approaches to explore the evolving impact of gaming across academic semesters or developmental stages.

Acknowledgments

We thank all respondents and UiTM Pahang for the opportunity to conduct this research.

References

[1] Hellström, C., Nilsson, K. W., Leppert, J., & Åslund, C. (2012). Influences of motives to play and time spent gaming on the negative consequences of adolescent online computer gaming. *Computers in human behaviour*, *28*(4), 1379-1387.

[2] Islam, M. I., Biswas, R. K., & Khanam, R. (2020). Effect of internet use and electronic game- play on academic performance of Australian children. *Scientific reports*, *10*(1), 21727.

[3] Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., Rothstein, H. R., & Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behaviour in eastern and western countries: a meta-analytic review. *Psychological bulletin*, *136*(2), 151.

[4] Granic, I., Lobel, A., & Engels, R. C. (2014). The benefits of playing video games. *American psychologist*, *69*(1), 66.

[5] Shin, H., & Ahn, J. (2023). *The influence of violent video games on aggression and stress: A meta-analytic review*. Current Psychology, 42,12654–12672. https://doi.org/10.1007/s12144-024-05667-5

[6] Yoo, C. (2024). *Psychosocial factors and changing patterns of prolonged online gaming in adolescents*. Current Psychology, 43(20), 18375-18388. https://doi.org/10.1007/s12144-024-06212-0

[7] Alfano, C. A., Zakem, A. H., Costa, N. M., Taylor, L. K., & Weems, C. F. (2009). Sleep problems and their relation to cognitive factors, anxiety, and depressive symptoms in children and adolescents. *Depression and anxiety*, *26*(6), 503-512.

[8] Kolhar, M., Kazi, R. N. A., & Alameen, A. (2021). Effect of social media use on learning, social interactions, and sleep duration among university students. *Saudi journal of biological sciences*, *28*(4), 2216-2222.

[9] Krejcie, R. V. ; M. D. W. (1970). Determining Sample Size for Research Activities. *Educational and Psychological Measurement*, *30*.

[10] Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of

Medical Education. PMCID: PMC4205511

[11] Taber, K. S. (2018). *The use of Cronbach's alpha when developing and reporting research instruments in science education*. *Research in Science Education, 48*(5), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

[12] Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Sage Publications.

Thermal Instability Analysis of Convection in a Horizontal Nanofluid Layer with Rigid-Rigid Boundary Condition

Norazuwin Najihah Mat Tahir[1,*], Norshahirah Izati Jaaffar[1], Nurul Nabila Rozali[1]
[1]*Center of Mathematical Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARACawangan Negeri Sembilan, Kampus Seremban, Persiaran Seremban Tiga/1, 70300 Seremban, Negeri Sembilan*
[*]najihahtahir@uitm.edu.my

Abstract

This study investigates the phenomenon of double-diffusive convection in viscoelastic nanofluids under the influence of external magnetic fields. The findings provide vital insights into how temperature gradients and magnetic forces affect fluid layer stability, demonstrating that these interactions can either enhance or delay the onset of convective instabilities. The study's goal is to develop a robust mathematical model that captures the dynamics of these interactions and employs linear stability analysis to understand how they affect fluid dynamics. The main objective is to transform the governing partial differential equations (PDE) into a system of ordinary differential equations (ODE) by using the linear stability analysis. The Galerkin-type weighted residual method is used to derive analytical solutions that explain the stability characteristics of the system. The influence of the scaled stress relaxation parameter, scaled strain retardation parameter and Chandrasekhar number on the system's stability was analysed with result interpretation carried out using Maple software. The stress relaxation advance the onset of oscillatory convection while the strain retardation delay the onset of oscillatory convection. The presence of magnetic field introduces a stabilizing effect on the nanofluid layer. These findings suggest that adjusting magnetic field strength and viscoelastic parameters can effectively regulate convective behavior in nanofluid systems, with potential applications in thermal management and energy-related technologies.

Keywords*:* Stress Relaxation, Strain Retardation, Chandrasekhar Number, Rigid-Rigid Boundary Condition

Introduction

Convective instabilities in fluid systems arise from variations in temperature, concentration, and surface tension, leading to buoyancy-driven motion that significantly affects heat and mass transfer efficiency. Among these phenomena, double-diffusive convection resulting from simultaneous thermal and solutal gradients has garnered increasing attention due to its critical role in geophysical, environmental, and engineering systems [1]. Nanofluids, which consist of nanoparticles suspended in a base fluid, exhibit enhanced thermal conductivity and heat transfer performance, making them highly suitable for such applications. When subject to external magnetic fields, these nanofluids demonstrate unique behaviours, particularly in electrically conductive environments, a phenomenon known as magnetoconvection [2]. Viscoelastic

nanofluids, which combine both viscous and elastic properties, add further complexity to the study of convection due to their non-Newtonian behaviour, which is influenced by factors such as stress relaxation and strain retardation [3].

The analysis of thermal instability in these complex fluids is vital for advancing technologies in fields such as nuclear cooling, biomedical transport, and energy storage. While previous studies have examined the influence of magnetic fields, nanoparticle concentration, and porous structures independently, limited research has integrated these factors with realistic boundary conditions such as rigid-rigid confinements. These rigid boundaries are especially relevant in practical settings, such as microfluidic devices and industrial reactors, where the fluid is mechanically constrained [4]. This study aims to fill this gap by employing linear stability analysis and the Galerkin-type weighted residuals method to analytically investigate the onset of stationary and oscillatory convection in a viscoelastic nanofluid layer. The findings will provide a deeper understanding of how physical parameters such as the Chandrasekhar number and viscoelastic relaxation times influence flow stability, contributing to improved designs in thermally driven fluid systems. Resulting from growing interests in heat transfer and magnetoconvective performance, research on thermal instability in nanofluids has been of immense academic interest. This review is intended to address present investigations forming the foundation of the influence of an external magnetic field, boundary conditions, and viscoelastic fluid behaviour on the onset of convection. Thermal instability normally occurs in nanofluid systems by way of buoyancy-induced convection due to temperature and solute concentration gradients. Central to such systems is the Rayleigh number, a measure of the balance between thermal diffusion and buoyancy, and whether a fluid system will be stable or convective.

A pivotal study by Yadav et al. [5] explored the role of external magnetic fields in Jeffrey nanofluid convection, showing that magnetic fields could significantly stabilise thermal and solutal instabilities. The research demonstrated that parameters like the Chandrasekhar number, Q and Lewis number, Le delay the onset of convection by counteracting the thermal gradients that drive instability. Their models incorporated Brownian motion and thermophoresis, which are critical components of nanofluid transport theory, yielding analytical solutions using Galerkin methods. These insights were further reinforced by Yadav et al. [6], who examined internal heating in nanofluid systems and highlighted that non-uniform heat sources and nanoparticle concentrations tend to destabilise the system, though this can be counteracted by increasing porosity and Darcy number.

Extending research on nanoparticle effetcs, Nield and Kuznetsov [7] studied the effect of nanoparticle dispersion and revealed that top-heavy configurations in particle concentration tend to stabilise the nanofluid layer, whereas bottom-heavy configuration promote oscillatory instabilities. This study emphasised that magnetic fields may either dampen or amplify convection depending on how they interact with the particle distribution and base thermal profiles. In a related study, Kumar et al. [3] adopted Jeffrey fluid dynamics to analyse viscoelastic nanofluids in porous media, showing that while magnetic fields stabilise, parameters like Jeffrey number and nanoparticle Rayleigh number promote convection, especially in biomedical and nuclear applications. Kuznetsov [8] expanded on Buongiorno's

theory by including both Brownian diffusion and thermophoresis in analysis of the oscillatory and stationary convection modes. Their stability analysis revealed that nanoparticle-induced slip mechanisms can be harnessed to improve bioconvection, a principle that is valuable in microfluidic devices for diagnostics and drug delivery systems.

Boundary conditions also significantly affect thermal instability. Sharma et al. [9] investigated convection in Jeffrey nanofluids under rigid-rigid boundaries, utilising the Galerkin-type weighted residual method. They found that rigid boundaries elevate the critical Rayleigh number, making the system more resistant to convection, which is essential in applications such as geothermal reservoirs and medical fluid transport. Complementing this, Ahuja and Gupta [10] showed that rigid-rigid boundaries in nanofluids enhanced system stability even under magnetic influence, suggesting stronger boundary resistance to fluid motion.

Mahajan and Sharma [4] used Chebyshev pseudospectral methods to reinforce that rigid-rigid conditions reduce flow disruptions, especially when exposed to gravity variations or magnetic fields. Their findings are applicable to systems with strict physical confinement, such as microgravity environments or nuclear cooling. Moreover, Yadav [11] confirmed that the no-slip conditions inherent to rigid boundaries significantly influence temperature gradients and flow behaviours, thus acting as a strong stabilising mechanism in nanofluid layers. The techniques employed to analyse stability have evolved from, classical linear stability analysis to the Galerkin methods and eigenvalue problems. For instance, Yadav et al. [6] used a linear stability analysis combined with eigenvalue formulations to determine critical convection thresholds. Narayana et al. [12] introduced a weakly nonlinear analysis and reduced their model to a generalised Lorenz system to explore the oscillatory patterns under thermal and solute gradients. Electro-thermal influences were the focus of Chand [13], who investigated nanofluids in Brinkman porous media. Using Galerkin methods and Newton's approximation, they showed how factors like Brinkman–Darcy number and electric Rayleigh number modify the critical Rayleigh number. Yadav et al. [14] extended this by demonstrating how the presence of a magnetic field modifies flow stability, revealing physical insights into dynamic instabilities.

Studies such as Ketchate et al. [15] and Boulal et al. [16] used harmonic balance methods and Galerkin projections to model periodic solutions in ferrofluid systems. These works illustrate how vibrational phenomena and external magnetic fields influence convection in ferrofluid layers, especially those used in targeted drug delivery or thermal therapy. Finally, Govender [17] examined the role of Darcy models in determining convection onset in vertically vibrated nanofluid layers and emphasised that critical Rayleigh numbers depend on both porosity and thermal properties, making them sensitive indicators of the stability of heating systems. Garg et al. [18] analysed triply diffusive thermo-bio-convection in anisotropic porous media and concluded that microorganism motility and internal heating contribute to both stabilising and destabilising influences, depending on system symmetry. Motivated by these insights, the current study aims to investigate the non-Newtonian behaviour of nanofluids, with particular emphasis on how parameters such as nanoparticle volume fraction, fluid viscosity and external forces for both surface and body affect convective instability. By conducting a theoretical analysis, this work seeks to enhance the understanding of how these variables govern

the onset of stationary and oscillatory instabilities and to uncover the underlying mechanisms that may either inhibit or amplify convective motion.

Methodology

Mathematical Model

This section provides a concise overview of the formulation and schematic representation of the case study. The study examines double-diffusive convection in a viscoelastic fluid of Oldroyd-B type with weak electrical conductivity. Figure 1 illustrates the system, which consists of a horizontal channel of infinite lateral extent and depth H, represented in a Cartesian coordinate system. The lower surface is placed at $z = 0$ and the upper surface at $z = H$. A uniform adverse temperature gradient, ΔT is maintained across the channel, driving convection. Gravity, $\boldsymbol{g} = -g\hat{e}_z$ acts vertically downwards, while an externally applied magnetic field of uniform strength H_0 is aligned along the positive z-direction.

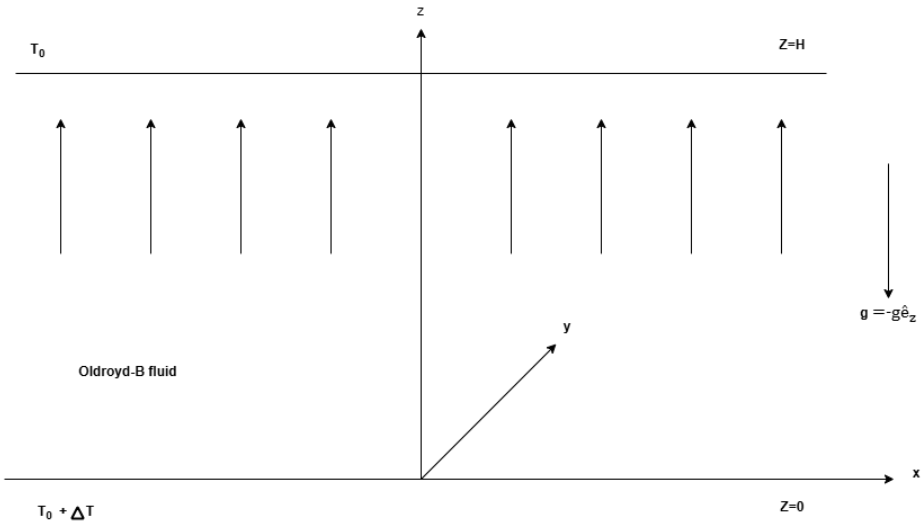


Figure 1: Schematic diagram of the physical problem

Governing Non-Linear Equation (PDE)

The governing equations for convective instabilities in a horizontal viscoelastic nanofluid layer are mathematical models that describe fluid behaviour. These equations are formulated as partial differential equations (PDEs) that involve various parameters. This study focuses on four key equations: conservation of mass, momentum, energy and nanoparticles.

$$\nabla \cdot \vec{v} = 0. \tag{1}$$

$$\left(1+\lambda_1\frac{\partial}{\partial t}\right)\left[\rho_{f_0}\left(\frac{\partial\vec{v}}{\partial t}+\vec{v}\cdot\nabla\vec{v}\right)+\nabla p-\rho\mathbf{g}+\sigma u\mu_m^2H_0^2\hat{e}_x+\sigma u\mu_m^2H_0^2\hat{e}_y\right] = \mu\left(1+\lambda_2\frac{\partial}{\partial t}\right)\nabla^2 v. \tag{2}$$

$$(\rho c)_f\left(\frac{\partial T}{\partial t}+\vec{v}\cdot\nabla\mathrm{T}\right)=\left(k\nabla^2T\right)+(\rho c)_p\left(D_B\nabla\phi\cdot\nabla T+\left(\frac{D_T}{T_C}\right)\nabla T\cdot\nabla T\right). \tag{3}$$

$$\frac{\partial\phi}{\partial t}+\vec{v}\cdot\nabla\phi=D_B\nabla^2\phi+\left(\frac{D_T}{T_C}\right)\nabla^2T, \tag{4}$$

where $\vec{v}=(u,v,w)$ is the nanofluid velocity, p is the pressure, λ_1 is the relaxation time, λ_2 is the retardation time, t is the time, σ is the electrical conductivity, μ_m is the magnetic permeability, μ is the coefficient of viscosity, ϕ is the nanoparticle volume fraction, T is the temperature, $(pc)_f$ is the effective heat capacity of the fluid, $(pc)_p$ is the effective heat capacity of the nanoparticle, k is the thermal conductivity, D_B is the Brownian diffusion coefficient and D_T is the thermophoretic diffusion coefficient. The nanofluid density, ρ is

$$\rho\cong\phi\rho_p+(1-\phi)\left[\rho_{f_0}-\rho_{f_0}\beta\left(T-T_c\right)\right], \tag{5}$$

where ρ_p is the nanoparticle mass density, ρ_{f_0} is the nanofluid density at the reference temperature T_c and β are the volumetric coefficient of thermal expansion. The temperature and the volumetric fraction of the nanoparticles are constant at the boundaries. In this study, the boundary conditions are rigid lower and upper boundaries will be considered. Therefore, the boundary conditions are

$$w=0,\quad \frac{\partial w}{\partial z}=0,\quad T=T_h,\quad \phi=\phi_0,\quad \text{at } z=0, \tag{6}$$

$$w=0,\quad \frac{\partial w}{\partial z}=0,\quad T=T_c,\quad \phi=\phi_1,\quad \text{at } z=H. \tag{7}$$

Linear Stability Analysis

The introduction of the scaling quantities for the length, velocity, time, pressure, nanoparticles volume fraction and temperature give the non-dimensional variables by

$$(x^*,y^*,z^*)=\frac{(x,y,z)}{H},(u^*,v^*,w^*)=(u,v,w)\frac{H}{\alpha_f},$$
$$t^*=\frac{\alpha_f}{H^2}t,p^*=\frac{H}{\mu\alpha_f}p,\phi^*=\frac{\phi-\phi_0}{\phi_1-\phi_0},T^*=\frac{T-T_c}{T_h-T_c}, \tag{8}$$

where $\alpha_f=\frac{k}{(\rho c)_f}$ is the thermal diffusivity of the fluid and asterisks denote dimensionless quantities. The resulting equations (1) – (7) in the form of the non-dimensional equations after dropping the asterisks by the scaling quantities (8) are

$$\nabla\cdot\vec{v}=0. \tag{9}$$

$$(1+\Lambda_1\frac{\partial}{\partial t})[\frac{1}{P_r}(\frac{\partial \vec{v}}{\partial t}+\vec{v}\cdot\nabla\vec{v})+\nabla p+Rm\hat{e}_z-RaT\hat{e}_z+Rn\phi\hat{e}_z+Q(u\hat{e}_x+\vec{v}\hat{e}_y)] = (1+\Lambda_1\frac{\partial}{\partial t})\nabla^2\vec{v}, \tag{10}$$

$$\frac{\partial T}{\partial t}+\vec{v}\cdot\nabla T=\nabla^2 T+\frac{N_B}{Le}\nabla\phi\cdot\nabla T+\frac{N_A N_B}{Le}\nabla T\cdot\nabla T, \tag{11}$$

$$\frac{\partial \phi}{\partial t}+\vec{v}\cdot\nabla\phi=\frac{1}{Le}\nabla^2\phi+\frac{N_A}{Le}\nabla^2 T, \tag{12}$$

subject to the dimensionless boundary conditions

$$w=0,\quad \frac{\partial w}{\partial z}=0,\quad T=1,\quad \phi=0,\quad \text{at } z=0, \tag{13}$$

$$w=0,\quad \frac{\partial w}{\partial z}=0,\quad T=0,\quad \phi=1,\quad \text{at } z=1. \tag{14}$$

The non-dimensionless parameters are

$$\Lambda_1=\lambda_1\frac{\alpha_f}{H^2},\Lambda_2=\lambda_2\frac{\alpha_f}{H^2},P_r=\frac{\mu}{\rho_{f_0}\alpha_f},Le=\frac{\alpha_f}{D_B},Q=\frac{\sigma{\mu^2}_m{H^2}_0H^2}{\mu},$$
$$N_A=\frac{D_T({T^*}_h-{T^*}_c)}{D_B{T^*}_c({\phi^*}_1-{\phi^*}_0)},N_B=\frac{(\rho c)_p}{(\rho c)_f}({\phi^*}_1-{\phi^*}_0)Ra=\frac{[\rho_{f_0}\beta({T^*}_h-{T^*}_c)]gH^3}{\mu\alpha_f}, \tag{15}$$
$$Rn=\frac{[(\rho_p-\rho_{f_0})({\phi^*}_1-{\phi^*}_0)]gH^3}{\mu\alpha_f},Rm=\frac{[\rho_p{\phi^*}_0+\rho_{f_0}(1-{\phi^*}_0)]gH^3}{\mu\alpha_f},$$

where $\Lambda_1$is the scaled stress relaxation parameter, Λ_2 is the scaled strain retardation parameter, P_r is the Prandtl number, Le is the Lewis number, Q is the Chandrasekhar number, N_A is the modified diffusivity ratio, N_B is the modified particle density increment, Ra is the thermal Rayleigh number, Rn is the concentration Rayleigh number and Rm is the basic density Rayleigh number.

Perturbation and Linearization

The fluid is at rest in the reference steady basic state varying in the z-direction. The basic state is perturbation by infinitesimal perturbations

$$\vec{v}=\vec{v}',p=p_b+p',T=T_b+T',\phi=\phi_b+\phi', \tag{16}$$

where prime denotes the perturbation quantity and $p_b=p_0,T_b=1-z,\phi_b=z$ where p_0 is the constant reference pressure. The system (9) – (12) and the boundary conditions (13) and (14) are perturbed with infinitesimal perturbations (16). The linearized perturbed system is,

$$\left(1+\Lambda_1\frac{\partial}{\partial t}\right)\left[\frac{1}{P_r}\frac{\partial}{\partial t}\nabla^2 w'-{\nabla^2}_H RaT'+{\nabla^2}_H Rn\phi'+Q{\nabla^2}_H w'\right]=\left(1+\Lambda_2\frac{\partial}{\partial t}\right)\nabla^4 w, \tag{17}$$

$$\frac{\partial T'}{\partial t}-w'=\nabla^2 T'+\frac{N_B}{Le}\left(\frac{\partial T'}{\partial z}-\frac{\partial \phi'}{\partial z}\right)-2\frac{N_A N_B}{Le}\frac{\partial T'}{\partial z}, \tag{18}$$

$$\frac{\partial \phi'}{\partial t}+w'=\frac{1}{Le}\nabla^2\phi'+\frac{N_A}{Le}\nabla^2 T', \tag{19}$$

subject to boundary conditions,

$$w' = 0, \quad \frac{\partial w}{\partial z} = 0, \quad T' = 0, \quad \phi' = 0 \quad \text{at } z = 0, \tag{20}$$

$$w' = 0, \quad \frac{\partial w}{\partial z} = 0, \quad T' = 0, \quad \phi' = 0 \quad \text{at } z = 1, \tag{21}$$

where ∇^4 is the three-dimensional biharmonic operator, ∇^2 is the three-dimensional Laplacian operator and $\nabla^2{}_H$ is the one-dimensional Laplacian operator in the horizontal plane.

Normal Modes

The non-dimensional system (17) – (21) constitutes a linear boundary-value problem that can be solved using the method of normal modes given by

$$(w', T', \phi') = [W(z), \Theta(z), \Phi(z)]e^{(st+ia_x+ia_y)}, \tag{22}$$

where $W(z), \Theta(z)\ and\ \Phi(z)$ are the amplitudes of the velocity, temperature and nanoparticle's volume fraction, respectively. $\alpha = (\alpha^2{}_x + \alpha^2{}_y)^{\frac{1}{2}}$ is the total wave number and s is a dimensionless complex growth rate. The system in the form of the normal modes are

$$(1+\Lambda_1 s)\left[\frac{s}{P_r}\left(D^2-\alpha^2\right)W + QD^2W + Ra\alpha^2\Theta - Rn\alpha^2\Phi\right] = (1+\Lambda_2 s)\left(D^2-\alpha^2\right)^2 W, \tag{23}$$

$$W + \left(D^2 + \frac{N_B}{Le}D - 2\frac{N_A N_B}{Le}D - \alpha^2 - s\right)\Theta - \frac{N_B}{Le}D\Phi = 0, \tag{24}$$

$$W - \frac{N_A}{Le}\left(D^2-\alpha^2\right)\Theta - \left[\frac{1}{Le}\left(D^2-\alpha^2\right) - s\right]\Phi = 0, \tag{25}$$

with the boundary conditions at $z = 0$ and $z = 1$,

$$W = 0, \quad DW = 0, \quad \Theta = 0, \quad \Phi = 0, \tag{26}$$

where $D \equiv \frac{d}{dz}$. The system with normal mode (23) – (26) is reduced to Nield and Kuznetsov [2] with $Q = 0$.

Analytical Computation using Galerkin-type Weighted Residuals Method

The Galerkin-type weighted residuals method was used to obtain an approximation of the closed-form solution to the system. The functions W, Θ and Φ are in the form of

$$W = \sum_{n=1}^{N} A_n W_n, \Theta = \sum_{n=1}^{N} B_n \Theta_n, \phi = \sum_{n=1}^{N} C_n \phi_n, \tag{27}$$

where A_n, B_n and C_n are unknown coefficients and $n = 1,2,3, \ldots, N$. Equations (23) – (26) are multiplied by W, Θ and Φ, respectively (27). Performing integration by parts with respect to z from 0 to 1 will lead to a system of $N-1$ equation with $N-1$ unknown. For rigid-rigid boundaries, the trial functions are

$$W_1 = z^2(1-z)^2, \quad \Theta_1 = z(1-z), \quad \phi_1 = z(1-z), \tag{28}$$

where W_1, Θ_1 and ϕ_1 are trial functions that satisfy the boundary conditions (26) [2]. The thermal Rayleigh number Ra determines the threshold for the transition from stationary patterns

to weakly chaotic evolution to a highly turbulent state. For $s = 0$, we obtain the solution for the case of stationary convection

$$Ra_{stat} = \frac{28(\alpha^4 + 24\alpha^2 + 504)(\alpha^2 + 10)}{27\alpha^2} + \frac{112Q(\alpha^2 + 10)}{9\alpha^2} - (Na + Le)Rn. \quad (29)$$

The stationary instability does not depend on the Λ_1 and Λ_2. For the case of Newtonian fluid, where $\Lambda_1 = \Lambda_2 = 0$ (no scaled stress relaxation and scaled strain retardation parameters) and in the absence of magnetic field, Eq. (29) reduces to the result of Nield and Kuznetsov [2] for stationary convection. Setting $s = i\omega$, we obtain the solution for $Ra = Ra_r + iRa_i$. Setting $Ra_i = 0$ will give the corresponding $\omega \neq 0$ for oscillatory convection and substitution of ω on Ra_rwill give the Rayleigh number for the oscillatory convection,

$$Ra_{osc} = \left(\frac{28}{27\alpha^2}\right)\{(\alpha^4 + 24\alpha^2 + 504)(\alpha^2 + 10)\left(\frac{1 + \Lambda_1\Lambda_2\omega^2}{1 + {\Lambda_1}^2\omega^2}\right)$$

$$-(\alpha^4 + 24\alpha^2 + 504)\omega^2\left(\frac{\Lambda_2 - \Lambda_1}{1 + {\Lambda_1}^2\omega^2}\right) - \left(\frac{12 + \alpha^2}{Pr}\right)\omega^2 \quad (30)$$

$$+12Q(10 + \alpha^2)\} - \left[\frac{(10 + \alpha^2)^2(Na + Le) + \omega^2 Le^2}{(10 + \alpha^2) + \omega^2 Le^2}\right]Rn.$$

For the case of Newtonian fluid, where $\Lambda_1 = \Lambda_2 = 0$ (no scaled stress relaxation and scaled strain retardation parameters) and in the absence of a magnetic field, Eq. (30) reduces to the result of Nield and Kuznetsov [2] for oscillatory convection.

Result and Discussion

The marginal stability curves for the onset of stationary and oscillatory instabilities were plotted using Maple software. Graphical illustrations of the marginal stability curves were analysed to assess the effects of the physical parameters. This section reports on the marginal stability curves derived using the Galerkin-type weighted residuals method. Our analysis elucidated the influence of key physical parameters, including the Chandrasekhar number, on the stability of double-diffusive convection in viscoelastic nanofluids. The results demonstrate the complex interplay between these parameters, which can either stabilise or destabilise the fluid layer.

Figure 2 presents the neutral stability curves for the variations of thermal Rayleigh number, Ra as a function of the wave number α for different values of the scaled stress relaxation parameter Λ_1, with values $\Lambda_1 = 0.2, 0.6,$ and 1.0. Stress relaxation refers to the process by which a fluid takes time to recover after a reduction in the applied deformation rate. The stress relaxation parameter does not depend on the stationary Rayleigh number. The oscillatory thermal Rayleigh number decreases with an increase in Λ_1. This trend indicates that the fluid becomes less resistant to convection as the stress relaxation parameter increases. The increase in Λ_1 allows the applied stress to influence the fluid element over a longer duration, resulting in reduced elastic memory. This reduced the internal friction, of the fluid. Consequently, convection initiates at lower values of the Darcy-Rayleigh number. Therefore, increasing Λ_1 advance the onset of oscillatory convection in the viscoelastic nanofluid layer, making the system more thermally unstable.

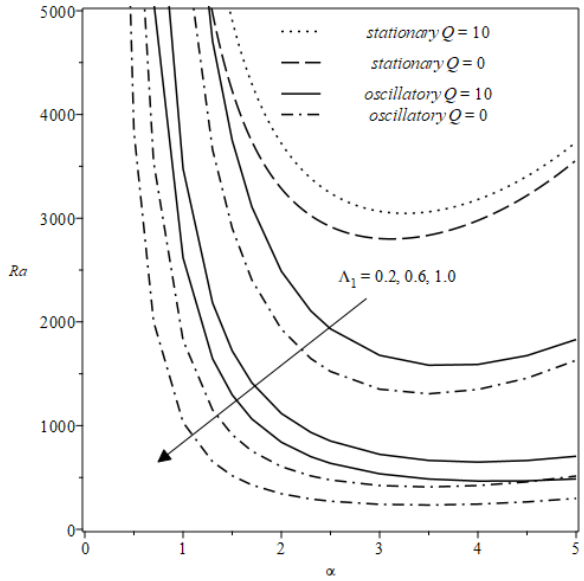


Figure 2: Neutral stability curve of the stationary and oscillatory convection for various values of Λ_1 when $\Lambda_2 = 0.1, P_r = 5, Le = 100,\ N_A = 5$ and $Rn = -10$

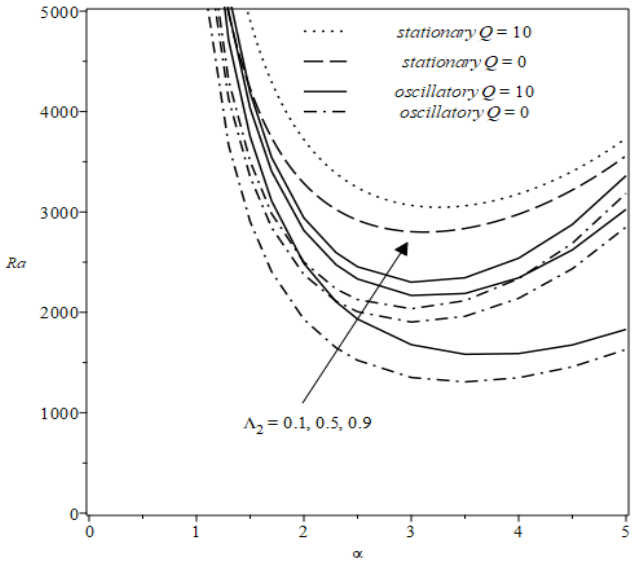


Figure 3: Neutral stability curve of the stationary and oscillatory convection for various values of Λ_2 when $\Lambda_1 = 0.2, P_r = 5, Le = 100,\ N_A = 5$ and $Rn = -10$

Figure 3 illustrates the neutral stability curves for the variations of thermal Rayleigh number, Ra as a function of the wave number α for various values of the scaled strain retardation parameter Λ_2, while other parameters are held constant. The strain retardation also does not depend on the stationary convection. This is because stress relaxation and strain retardation do not depend on time during the perturbation solution. It has been discovered that an increase in

the value of Λ_2 leads to a higher critical Rayleigh number for the onset of oscillatory convection, indicating that the system becomes thermally more stable. Parameter Λ_2 represents the time taken by the fluid to respond to an applied stress. A higher value of Λ_2 means that the fluid has a stronger resistance to deformation, which increases internal friction and delays the onset of convection. This stabilising effect is particularly significant in the oscillatory mode, as indicated by the upward shift of the neutral curves. In contrast, the stationary mode remains mostly unaffected by changes in Λ_2. This behaviour occurs because the stationary instability is independent of the viscoelastic parameters, as the base state remains equivalent to pure conduction, similar to a Newtonian nanofluid.

Figure 4 illustrates the neutral stability curves, plotted against the Rayleigh number Ra, and wave number α, for different values of the magnetic parameter Q. As shown in the figure, the critical Rayleigh number increases with increasing values of Q, for both stationary and oscillatory convection, indicating that the presence of a magnetic field introduces a stabilising effect on the nanofluid layer. This is due to the Lorentz force, represented by Q, which resists fluid motion and acts as a damping mechanism. As Q increases, the damping becomes stronger, requiring more thermal energy to initiate convection. The influence of Q is more prominent in the oscillatory mode, where the curves shift upward significantly, reflecting the greater sensitivity of dynamic fluid motion to electromagnetic resistance. In contrast, the stationary mode exhibits a more gradual rise in critical Ra with increasing Q, indicating a slower onset of convection. Nonetheless, the presence of a magnetic field still contributes to thermal stabilisation even in the absence of oscillatory motion. Overall, the increasing critical Rayleigh number for both modes confirmed that the magnetic fields suppressed convection by enhancing the system stability through electromagnetic damping.

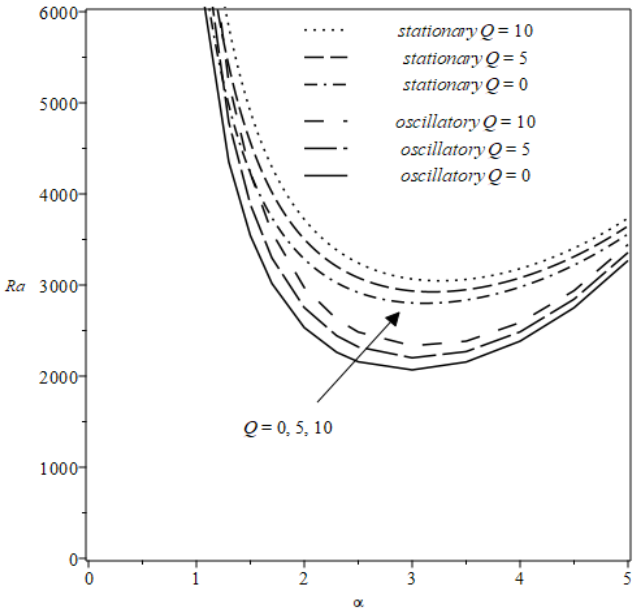


Figure 4: Neutral stability curve of the stationary and oscillatory convection for various values of Q when $\Lambda_1 = 0.2,\ \Lambda_2 = 1.0,\ P_r = 5, Le = 100,\ N_A = 5$ and $Rn = -10$.

Conclusion

This study discusses the thermal instability of convection in a rigid-rigid horizontal nanofluid layer under the effects of viscoelastic parameters and magnetic field on the onset of convection. Linear stability and the Galerkin-type weighted residual method were used to simplify the governing partial differential equations into ordinary differential equations, for which the stability behaviour was analysed. The marginal stability curves obtained from the resulting equations proved that increases in the scaled stress relaxation parameter lowered the critical Rayleigh number, thus indicating an earlier convection onset and thus an increasingly thermally unstable system. Conversely, increases in the scaled strain retardation parameter and Chandrasekhar number increased the critical Rayleigh number, thereby stabilising the fluid system. Graphical analysis also confirmed that the stationary convection was largely insensitive to viscoelastic parameters, while oscillatory modes were highly sensitive to changes in Λ_1, Λ_2, and Q. The findings locate the complex dynamic interaction of viscoelastic and magnetic effects in controlling the convective stability in nanofluids. Furthermore, the findings coincide with previous theoretical expectations and validate the appropriateness of the Galerkin method for such fluid dynamics problems. Finally, the study demonstrates that magnetic fields are beneficial for delaying the onset of thermal convection and enhancing the thermal stability of nanofluids. This is particularly valuable in applications such as biomedical cooling, electronics cooling, and energy systems. However, this study also highlights the importance of selecting parameters and boundary conditions for predicting system behaviour, thereby providing a foundation for future experimental and computational investigations of nanofluid convection dynamics.

Acknowledgments

The authors express their sincere gratitude to everyone who contributed to the success of this study, especially for their unwavering support and encouragement throughout this journey.

References

[1] Yadav, D., Mohamed, R. A., Lee, J. and Cho, H. H. (2017). Thermal convection in a Kuvshiniski viscoelastic nanofluid saturated porous layer. *Ain Shams Engineering Journal,* 8(4), 613–621. https://doi.org/10.1016/j.asej.2015.11.023

[2] Nield, D. A. and Kuznetsov, A. V. (2010). The onset of convection in a horizontal nanofluid layer of finite depth. *European Journal of Mechanics-B/Fluids*, 29(3), 217–223. https://doi.org/10.1016/j.euromechflu.2010.02.003

[3] Kumar, A. K. A., Lata, P., Lata, P. L. P., Bains, D. B. D. and Thakur, P. T. P. (2024). Effect of magnetic field on the onset of thermal convection in a Jeffery nanofluid layer saturated by a porous medium: free-free, rigid-rigid and rigid-free boundary conditions. *Journal of the Nigerian Society of Physical Sciences*, 1934–1934. https://doi.org/10.46481/jnsps.2024.1934

[4] Mahajan, A. and Sharma, M. K. (2018). The onset of convection in a magnetic nanofluid layer with variable gravity effects. *Applied Mathematics and Computation*, 339, 622–635. https://doi.org/10.1016/j.amc.2018.07.062

[5] Yadav, D., Bhargava, R. and Agrawal, G. S. (2013). Numerical solution of a thermal instability problem in a rotating nanofluid layer. *International Journal of Heat and Mass Transfer*, 63, 313–322. https://doi.org/10.1016/j.ijheatmasstransfer.2013.04.003

[6] Yadav, D., Bhargava, R. and Agrawal, G. S. (2012). Boundary and internal heat source effects on the onset of Darcy-Brinkman convection in a porous layer saturated by nanofluid. *International Journal of Thermal Sciences*, 60, 244–254. https://doi.org/10.1016/j.ijthermalsci.2012.05.011

[7] Nield, D. A. and Kuznetsov, A. V. (2009). Thermal instability in a porous medium layer saturated by a nanofluid. *International Journal of Heat and Mass Transfer*, 52(25–26), 5796–5801. https://doi.org/10.1016/j.ijheatmasstransfer.2009.07.023

[8] Kuznetsov, A. V. (2011). Non-oscillatory and oscillatory nanofluid bio-thermal convection in a horizontal layer of finite depth. *European Journal of Mechanics-B/Fluids*, 30(2), 156–165. https://doi.org/10.1016/j.euromechflu.2010.10.007

[9] Sharma, P. L., Kumar, A., Bains, D., Lata, P. and Rana, G. C. (2023). Thermal convective instability in a Jeffrey nanofluid saturating a porous medium: rigid-rigid and rigid-free boundary conditions. *Structural Integrity and Life*, 23(3), 351–356. https://doi.org/10.5281/zenodo.15720378

[10] Ahuja, J. and Gupta, U. (2019). Rayleigh-Bénard Convection for Nanofluids for More Realistic Boundary Conditions (Rigid-Free and Rigid-Rigid) Using Darcy Model. *International Journal of Mathematical, Engineering and Management Sciences*, 4(1), 139–156. https://doi.org/10.5281/zenodo.15721059

[11] Yadav, D. (2017). Electrohydrodynamic instability in a heat generating porous layer saturated by a dielectric nanofluid. *Journal of Applied Fluid Mechanics*, 10(3), 763–776. https://doi.org/10.18869/acadpub.jafm.73.240.27475

[12] Narayana, M., Gaikwad, S. N., Sibanda, P. and Malge, R. B. (2013). Double diffusive magneto-convection in viscoelastic fluids. *International Journal of Heat and Mass Transfer*, 67, 194–201. https://doi.org/10.1016/j.ijheatmasstransfer.2013.08.027

[13] Chand, R. (2017). Electro-thermal convection in a Brinkman porous medium saturated by nanofluid. *Ain Shams Engineering Journal*, 8(4), 633–641. https://doi.org/10.1016/j.asej.2015.10.008

[14] Yadav, D., Wang, J., Bhargava, R., Lee, J. and Cho, H. H. (2016). Numerical investigation of the effect of magnetic field on the onset of nanofluid convection. *Applied Thermal Engineering*, 103, 1441–1449. https://doi.org/10.1016/j.applthermaleng.2016.05.039

[15] Ketchate, C. G. N., Fokwa, D., Tchuen, G. and Kapen, P. T. (2021). Stability analysis of non-Newtonian blood flow conveying hybrid magnetic nanoparticles as target drug delivery in presence of inclined magnetic field and thermal radiation: Application to therapy of cancer. *Informatics in Medicine Unlocked*, 27. https://doi.org/10.1016/j.imu.2021.100800

[16] Boulal, T., Riahi, M. and Aniss, S. (2024). Thermo-magnetic convection and convective magneto-gravitational parametric resonance in a horizontal ferrofluid layer under time quasi-periodic vibrations: A balance harmonic analysis. *Chinese Journal of Physics*, 89, 1286–1298. https://doi.org/10.1016/j.cjph.2024.01.003

[17] Govender, S. (2017). Thermal instability in a nanofluid saturated horizontal porous layer subjected to g-gitter. *International Journal of Heat and Mass Transfer*, 110, 63–67. https://doi.org/10.1016/j.ijheatmasstransfer.2017.03.018

[18] Garg, A., Sharma, Y. D. and Jain, S. K. (2023). Onset of triply diffusive thermo-bio-convection in the presence of gyrotactic microorganisms and internal heating into an anisotropic porous medium: oscillatory convection. *Chinese Journal of Physics*, 84, 173–188. https://doi.org/10.1016/j.cjph.2023.05.014

PREDICTING GOLD MARKET TRENDS USING NEWTON'S DIVIDED DIFFERENCE METHOD

Nurul Athirah Soleha Mohd Zin[1], Nurshazrin Izzati Azat[1], Nor Hanim Abd Rahman[2], Rafizah Kechil[2]*

[1]Universiti Teknologi MARA Cawangan Negeri Sembilan Kampus Seremban

[2]Universiti Teknologi MARA Cawangan Pulau Pinang Kampus Permatang Pauh

*norhanim@uitm.edu.my

Abstract

Gold is a vital financial asset often used as a hedge in times of economic instability. This study explores the application of Newton's Divided Difference (NDD) method to forecast gold market trends using historical price data from March 2022 to April 2025. Both monthly (36 data points) and quarterly (12 data points) datasets were analyzed to assess the impact of data resolution on model accuracy and efficiency. Polynomial interpolation models of degree 3 and 7 were constructed using Newton's Forward and Backward Divided Difference techniques. Error metrics such as Absolute Error (AE) and Percentage Absolute Relative Error (PARE) were used to evaluate prediction accuracy. The results revealed that while monthly data offered finer detail, quarterly data produced more stable and interpretable forecasts. This study confirms that NDD is a practical method for financial trend prediction, offering a simple yet effective alternative to more complex algorithms.

Keywords*:* Gold price prediction, Newton's Divided Difference, interpolation, polynomial modeling, data frequency comparison

Introduction

Gold is a widely traded commodity and a safe-haven asset, especially during periods of economic uncertainty. Accurate forecasting of gold prices is essential for investors, financial analysts, and policymakers. Traditional prediction methods often rely on large datasets and complex machine learning models. This study aims to investigate a simpler yet powerful classical technique which is Newton's Divided Difference (NDD) interpolation for modeling and predicting gold price trends. Newton's Divided Difference is a classic numerical interpolation technique effective for handling unevenly spaced data points. Prior studies highlight its accuracy and computational simplicity, particularly in fields where data is limited but time-sensitive, such as financial forecasting. This work builds on such findings by evaluating NDD's effectiveness in predicting real-world gold price trends.

Background

Gold is a principal safe-haven asset, especially in times of economic and geopolitical uncertainty. Forecasting gold prices is crucial for informed financial decision-making [1]. Traditional forecasting models rely on complex algorithms and big data, making them less accessible [2],[3]. This study investigates NDD, a venerable numerical interpolation method valued for its ease of use and appropriateness to handle unevenly spaced data [3],[4]. Gold prices between March 2022 and April 2025 were analyzed in terms of datasets by month and

quarter. Order 3 and 7 polynomial models were constructed by way of forward and backward NDD methodologies. The purpose is to analyze the practicability and validity of NDD in predicting actual gold prices.

Methodology

The methodology of this study uses the historical gold price data from March 2022 to April 2025 collected from a reputable financial source and organized into two distinct datasets: monthly data containing 36 data points and quarterly data comprising 12 data points. Figure 1 below shows the source of the data from MSN Finance [5].



Figure 1: Gold Price from March 2022 – April 2025

For each dataset, a Newton's Divided Difference (NDD) table was created to build the interpolation polynomials. These tables display the gold price values along with their divided differences, calculated step by step up to the highest order. Table 1 illustrates this pyramid structure of the NDD table using the 12 data points.

Table 1: NDD Table

x_i	$f(x_i)$	$1^{st}DD$	$2^{nd}DD$	$3^{rd}DD$	$4^{th}DD$	$5^{th}DD$	$6^{th}DD$	$7^{th}DD$	$8^{th}DD$
0	1943.80								
		-31.7000							
3	1848.70		-2.4611						
		-46.4667		1.7006					
6	1709.30		12.8444		-0.2803				
		30.6000		-1.6636		0.0316			
9	1801.10		-2.1278		0.1943		-0.0031		
		17.8333		0.6679		-0.0237		0.0003	
12	1854.60		3.8833		-0.1607		0.0028		0.0000
		41.1333		-1.2605		0.0269		-0.0003	
15	1978.00		-7.4611		0.2428		-0.0034		0.0000
		-3.6333		1.6531		-0.0339		0.0004	
18	1967.10		7.4167		-0.2664		0.0041		0.0000
		40.8667		-1.5438		0.0395		-0.0004	
21	2089.70		-6.4778		0.3263		-0.0049		0.0000
		2.0000		2.3716		-0.0482		0.0005	
24	2095.70		14.8667		-0.3970		0.0051		
		91.2000		-2.3920		0.0432			
27	2369.30		-6.6611		0.2516				
		51.2333		0.6278					
30	2523.00		-1.0111						
		45.1667							
33	2658.50								

Newton's Forward Divided Difference

Newton's Backward Divided Difference

A similar derivation process was also carried out for the full dataset comprising 36 data points to obtain a more comprehensive interpolation. The resulting polynomial formulas derived from both datasets are presented as follows.

Polynomial equations of degree 3 and degree 7 (Monthly Data/All data):

NFDD:

$$P_3(x) = 1943.80 + 28.3x - 62.6x^2 + 14.2x^3 \quad (1)$$

$$P_7(x) = 1943.80 + 420.424x - 958.7658x^2 + 759.9459x^3 - 296.6702x^4 + 61.0432x^5 - 6.338 + 0.2613x^7 \quad (2)$$

NBDD:

$$P_3(x) = 2857.10 - 302.3666x + 126.9999x^2 - 12.7333x^3 \quad (3)$$

$$P_7(x) = 2857.10 - 68.9816x - 367.9796x^2 + 348.6684x^3 - 115.1657x^4 + 16.2047x^5 - 0.849 + 0.003x^7 \quad (4)$$

Polynomial equations of degree 3 and degree 7 (Quarterly Data):

NFDD:

$$P_3(x) = 1943.80 + 6.2941x - 17.7665x^2 + 1.7006x^3 \quad (5)$$

$$P_7(x) = 1943.80 + 360.9931x - 285.5732x^2 + 74.9957x^3 - 9.5533x^4 + 0.6436x^5 - 0.022x^6 + 0.0003x^7 \quad (6)$$

NBDD:

$$P_3(x) = 2658.50 - 53.4338x + 4.6391x^2 - 0.6278x^3 \quad (7)$$

$$P_7(x) = 2658.50 - 589.3298x + 415.3829x^2 - 115.5191x^3 + 15.3716x^4 - 1.0602x^5 + 0.0366x - 0.0005x^7 \quad (8)$$

To model and forecast future prices, both Newton's Forward Divided Difference (NFDD) and Backward Divided Difference (NBDD) interpolation techniques were applied to each dataset. Polynomial models of degree 3 and 7 were constructed using Maple software. The performance of each model was assessed using Absolute Error (AE) and Percentage Absolute Relative Error (PARE) as evaluation metrics.

Results

Table 2 below shows the results of comparison of both NFDD and NBDD methods by using monthly data.

Table 2: Comparison of NFDD and NBDD for Monthly Data

x_i	DEGREE 3				DEGREE 7			
	AE		PARE		AE		PARE	
	NFDD	NBDD	NFDD	NBDD	NFDD	NBDD	NFDD	NBDD
1	0	745.30	0	38.7430	0	745.30	0	38.7430
5	507.60	1140.9020	28.3940	63.8195	0.0012	735.3020	0.0001	41.1312
15	34230.30	18056.3090	1730.5511	912.8569	6138335.5462	1591355.3090	310330.4118	80452.7457
19	75433.50	46225.8062	4083.6672	2502.4798	52734161.6746	9895439.4830	2854816.0283	535699.4090
22	121396.20	79984.5920	5854.9339	3857.6537	185862928.0695	28730551.4600	8964161.6702	1385673.3607

A comparison between degree 3 and degree 7 polynomial models using NFDD and NBDD approaches reveals significant differences in performance. Degree 3 model shows reliable and consistent error growth on both approaches, which implies enhanced reliability for interpolation of monthly data. Degree 7 model produces very low error at interpolation points but a sharp increase at boundary values, especially with NFDD. This is Runge's phenomenon, and it indicates the instability of higher-degree polynomials. Generally, the degree 3 model is easier and reliable for practical use. Table 3 below shows the comparison of NFDD and NBDD methods using quarterly data for both degree 3 and degree 7 polynomial interpolation models.

Table 3: Comparison of NFDD And NBDD for Quarterly Data

x_i	DEGREE 3				DEGREE 7			
	AE		PARE		AE		PARE	
	NFDD	NBDD	NFDD	NBDD	NFDD	NBDD	NFDD	NBDD
1	10.3282	685.3775	0.5369	35.6281	161.5842	459.6815	8.3997	23.8957
5	44.0170	641.1335	2.4622	35.8636	83.6770	695.7735	4.6807	38.9200
15	1802.2740	1196.0345	91.1160	60.4669	0.7860	10.1945	0.0397	0.5154
19	5466.8968	2835.3073	295.9559	153.4922	32.8968	247.2527	1.7809	13.3853
22	9517.8730	5029.9336	459.0466	242.5935	1058.2218	1847.7832	51.0380	89.1185

A performance comparison between degree 3 and degree 7 polynomial interpolation models using NFDD and NBDD reveals distinct differences. The degree 3 model, particularly when applied with NFDD, yields more stable and consistent results with gradual error progression. In contrast, the degree 7 model achieves higher accuracy at central data points but exhibits significant boundary errors, especially under the NBDD method. These findings suggest that degree 3 with NFDD offers a more robust and balanced approach for gold price forecasting

based on quarterly data. Figure 2 shows the comparison between actual monthly gold prices and the forecasts generated using degree 3 and degree 7 polynomial models.

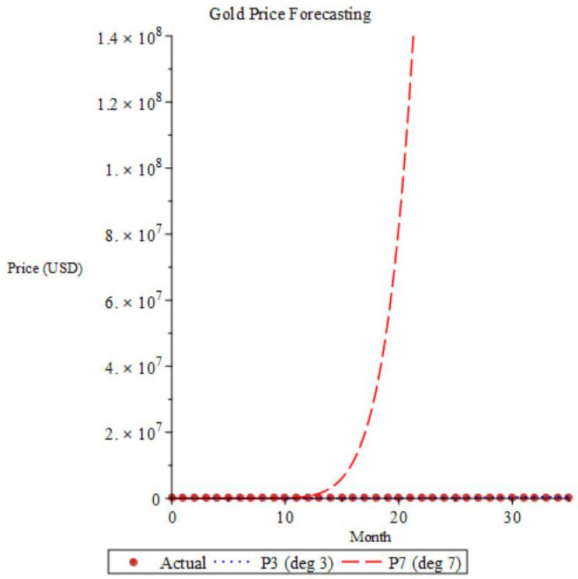


Figure 2: Graph Gold Price Forecasting (Monthly Data)

The graph is a comparison of real gold prices and those that are forecast by degree 3 and degree 7 polynomial models. The degree 3 model follows the real trend quite closely with smooth and stable outcomes over the course of the period. In contrast, the degree 7 model starts to deviate significantly after a certain point, showing sharp increases that do not match the real data This indicates that higher-order models may lead to overfitting and less precise predictions, especially near the extremes of the data set. Figure 3 below shows the comparison between actual quarterly gold prices with the predicted values from degree 3 and degree 7 polynomial models.

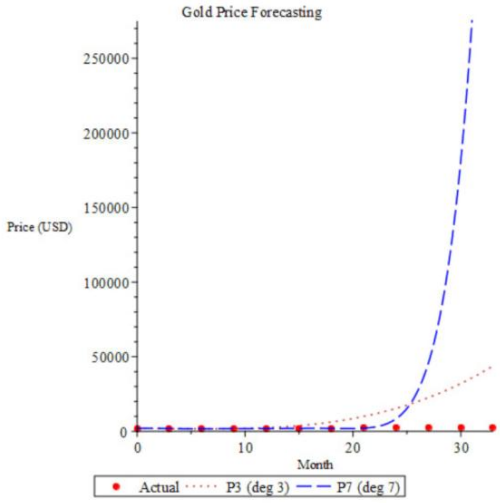


Figure 3: Graph Gold Price Forecasting (Quarterly Data)

This graph compares actual quarterly gold prices with predicted values generated by degree 3 and degree 7 polynomial models using NDD method. Both models initially follow the actual trend closely, but as the time progresses, the degree 3 model maintains a stable and realistic curve, while the degree 7 model shows a sharp increase and deviates significantly from the actual data. This highlights the risk of overfitting and instability in higher-degree models when applied to lower-frequency datasets.

Acknowledgments

The authors would like to express their gratitude to Dr. Nor Hanim binti Abd Rahman for her valuable supervision, direction, and assistance during this work. The authors would also like to express their gratitude to Puan Rafizah Kechil for technical assistance, particularly in the applying of the Newton's Divided Difference through Maple software. Appreciation is also extended to the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA for providing facilities and resources.

References

[1] D. K. Kushwaha, D. K. Sharma, S. S. Khullar, S. Shukla, T. K. Pandey, and S. Pal, "Gold price prediction using an ensemble of Random Forest and XGBoost," *Rivista Italiana di Analisi dei Dati*, vol. 14, no. 1, 2023. [Online]. Available: https://rifanalitica.it

[2] A. E. S. H. Maharani, D. A. Azka, and D. Darlena, "Gold price fluctuation forecasting based on Newton and Lagrange polynomial interpolation," Jurnal Varian, vol. 7, no. 1, pp. 87–98, Oct. 2023, doi: 10.30812/varian.v7i1.3230.

[3] L. Zou, L. Song, X. Wang, T. Weise, Y. Chen, and C. Zhang, "A New Approach to Newton-Type Polynomial Interpolation with Parameters," Mathematical Problems in Engineering, vol. 2020, pp. 1–15, Nov. 2020, doi: 10.1155/2020/9020541.

[4] B. Das and D. Chakrabarty, "Newton's Divided Difference Interpolation formula: Representation of Numerical Data by a Polynomial curve," *International Journal of Mathematics Trends and Technology*, vol. 35, 2016. [Online]. Available: http://www.ijmttjournal.org

[5] MSN Finance. (n.d.). Gold Price Chart (March 2022 – April 2025). Retrieved from https://www.msn.com/en-my/money/chart?id=auvwoc&timeFrame=3Y&chartType=line&projection=false

APPLICATION OF THE RUNGE-KUTTA FOURTH ORDER (RK4) AND RUNGE-KUTTA FIFTH ORDER (RK5) IN DYNAMIC SYSTEMS: A FOCUS ON ELECTRICAL CIRCUIT ANALYSIS

Nur Nayli Afiqah Abd Razak[1], Faranissha Shazlyn Hasmin Farouk[1], Nor Hanim Abd Rahman[2], Mohd Fahmi Zahari[2]*

[1]Universiti Teknologi MARA Cawangan Negeri Sembilan Kampus Seremban

[2]Universiti Teknologi MARA Cawangan Pulau Pinang Kampus Permatang Pauh

[*]norhanim@uitm.edu.my

Abstract

This study examines the application of the Runge-Kutta Fourth Order (RK4) and Runge-Kutta Fifth Order (RK5) numerical methods in analyzing dynamic electrical systems, with a particular focus on both series and parallel RLC circuits. These numerical techniques are renowned for their ability to solve ordinary differential equations (ODEs) with a high level of accuracy and computational efficiency, especially in scenarios where analytical solutions are difficult or impossible to derive. The research aims to compare the effectiveness of RK4 and RK5 in simulating the transient and steady-state behaviours of electrical circuits under various damping conditions, including underdamped, overdamped, and critically damped responses. The study employs Mathematica as the primary tool to compare the accuracy of RK4 and RK5 in simulating both the transient (short-term) and steady-state (long-term) responses of electrical circuits under various damping conditions. The performance of each method is evaluated based on accuracy, stability, and computational efficiency. The results show that RK4 is a solid choice for balancing simplicity and performance, while RK5 stands out for its higher accuracy, especially important in situations where precision matters most. By applying both methods to series and parallel circuit setups, the study gives a more complete picture of how these numerical approaches perform in real-world circuit analysis. Overall, this work helps engineers, researchers, and students make better decisions about which method to use based on the type of circuit and how precise the simulation needs to be. It highlights the trade-offs between RK4 and RK5 and shows how choosing the right method can improve the accuracy and efficiency of electrical circuit modelling.

Keywords*:* Runge-Kutta, RK4, RK5, Mathematica, Circuit Analysis

Introduction

Numerical methods help solve ordinary differential equations (ODEs), especially when it is hard to find exact solutions. In electrical engineering, these methods are used to study how circuits behave over time, particularly during sudden changes. Although Euler's method is simple to use, it often lacks the accuracy required for analyzing complex or sensitive systems [1].

Runge-Kutta methods, such as RK4 and RK5, are widely recognized for offering improved accuracy and stability. These methods approximate slopes at multiple points within a step, making them highly suitable for studying how circuits respond in real-time [2],[3]. This study

focuses on employing RK4 and RK5 to simulate the behavior of RLC circuits over time. Although both methods are commonly applied, there is limited research that directly compares them in detail for electrical circuit applications. Such a comparison is significant because electrical circuits, especially RLC systems, are susceptible to numerical precision when modeling transient oscillations, damping effects, and steady-state responses. Small differences in accuracy or stability between RK4 and RK5 can impact on how reliably simulations capture circuit behavior, which in turn affects design and performance decisions in areas such as power electronics, signal processing, and control systems.

Therefore, by analyzing RK4 and RK5 side by side using Mathematica software, this study aims to evaluate their accuracy and efficiency. The goal is to provide engineers and researchers with practical guidance on selecting the appropriate method based on circuit type, computational constraints, and precision requirements [5-6].

Methodology

To evaluate the performance of the Runge-Kutta Fourth Order (RK4) and Fifth Order (RK5) methods in electrical circuit analysis, this study applies both techniques to the governing differential equations of RLC circuits. Since RLC systems are modeled using second-order ordinary differential equations (ODEs), analytical solutions are often difficult to obtain, particularly for nonlinear or time-varying cases. Therefore, provide a practical way to approximate circuit behavior over time.

The implementation was carried out in Mathematica, using specified parameters for resistance (R), inductance (L), and capacitance (C). Simulations were performed for series and parallel RLC circuits under different damping conditions to capture both transient and steady-state responses. The performance of RK4 and RK5 was then compared in terms of accuracy and computational effort.

Equation (1) is the formula of the Runge-Kutta fourth-order method given by [5] as below:

$$y'' = f(x, y) \tag{1}$$

With initial conditions

$$y(x_0) = y_0 \ \ and \ y'(x_0) = y'_0$$

The RK4 formula is

$$y_{n+1} = y_n + \frac{1}{6}(k1 + 2(k2 + k3) + k4) + O(h^5)$$

where,

$$k1 = hf(t_n, y_n)$$

$$k2 = hf(t_n + \frac{1}{2}h, y_n + \frac{1}{2}R1)$$

$$k3 = hf(t_n + \frac{1}{2}h, y_n + \frac{1}{2}R2)$$

$$k4 = hf(t_n + h, y_n + R3)$$

h represents the step size, $\mathbf{t_n}$ are the values of the independent variables and $\mathbf{y_n}$ are the values of the dependent variables at the current step. k1, k2, k3, and k4 represent successive slope estimates that are combined to approximate the solution with fourth-order accuracy.

Although several RK5 formulations exist in the literature, in this study, RK5 implemented the formulation given by [4], shown in Eq.2. This specific scheme was chosen because it is well

established in electrical circuit applications and provides higher-order accuracy with manageable computational effort. The alternative RK5 formulation Eq. 3 was reviewed for comparison but was not implemented in the simulations.

$$y_{n+1} = y_n + \frac{h}{90}(7k_1 + 32k_3 + 12k_4 + 32k_5 + 7k_6) \tag{2}$$

where,

$$t_{n+1} = t_n + h,$$
$$k_1 = f(t_n, y_n),$$
$$k_2 = f\left(t_n + \frac{h}{4}, y_n + \frac{h}{4}k_1\right),$$
$$k_3 = f\left(t_n + \frac{h}{4}, y_n + \frac{h}{8}k_1 + \frac{h}{8}k_2\right),$$
$$k_4 = f\left(t_n + \frac{h}{2}, y_n - \frac{h}{2}k_2 + hk_3\right),$$
$$k_5 = f\left(t_n + \frac{3h}{4}, y_n + \frac{3h}{16}k_1 + \frac{9h}{16}k_4\right),$$
$$k_6 = f\left(t_n + h, y_n - \frac{3h}{7}k_1 + \frac{2h}{7}k_2 + \frac{12h}{7}k_3 - \frac{12h}{7}k_4 + \frac{8h}{7}k_5\right)$$

Also, the RK5 formula constructed by [7] as Eq. 3:

$$y_{n+1} = y_n + \frac{1}{192}(23k_1 + 125k_3 - 81k_5 + 125k_6) \tag{3}$$

where,

$$k_1 = hf(t_n, y_n)$$
$$k_2 = hf\left(t_n + \frac{h}{3}, y_n + \frac{k_1}{3}\right)$$
$$k_3 = hf\left(t_n + \frac{2h}{5}, y_n + \frac{1}{25}(4k_1 + 6k_2)\right)$$
$$k_4 = hf\left(t_n + h, y_n + \frac{1}{4}(k_1 - 12k_2 + 15k_3)\right)$$
$$k_5 = hf\left(t_n + \frac{2h}{3}, y_n + \frac{1}{81}(6k_1 + 90k_2 - 50k_3 + 8k_4)\right)$$
$$k_6 = hf\left(t_n + \frac{4h}{5}, y_n + \frac{1}{75}(6k_1 + 36k_2 + 10k_3 + 8k_4)\right)$$

h is the step size, t_n are the values of the independent variables and y_n are the values of the dependent variables. k1, k2, k3, k4, k5, and k6 are intermediate calculations based on the derivative function.

This requires the use of numerical techniques to solve the governing differential equations of the circuit. Parameters for the circuit are given as resistance (R), inductance (L), and capacitance (C).

Classification of circuit response:

The damping behavior of RLC circuits was classified based on the damping factor (ξ) using Eq. (4):

$$\xi = \frac{R}{2}\sqrt{\frac{C}{L}} \tag{4}$$

The circuit response is categorized as:

Overdamped ($\xi > 1$) : slow, non-oscillatory return to steady state.

Critically damped ($\xi = 1$): fastest return to steady state without oscillation.

Underdamped ($\xi < 1$) : oscillatory behavior with decaying amplitude.

This classification is crucial as it determines the nature of the transient response, ranging from convergence to steady state in overdamped systems to oscillatory behaviour in systems.

The accuracy of RK4 and RK5 was evaluated using the absolute error formula in Eq. (5) from [6]:

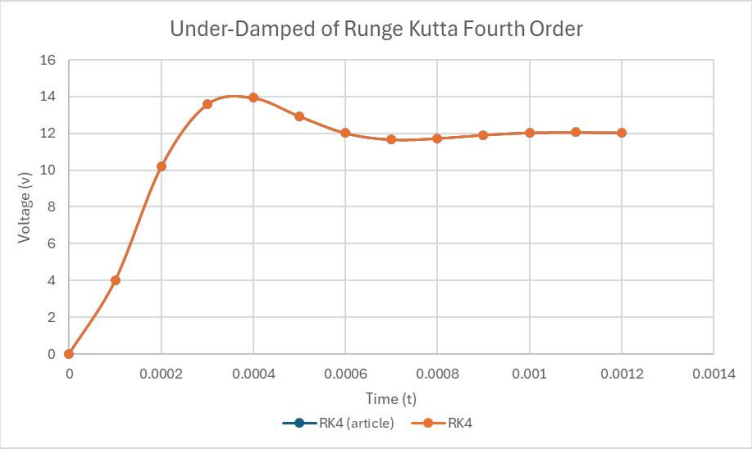
$$\texttt{Absolute Error} = |\boldsymbol{V_A} - \boldsymbol{V_E}| \tag{5}$$

where $\mathbf{V_A}$ is the analytical (exact) solution and $\mathbf{V_E}$ is the numerical approximation. This allows for a direct performance comparison of the two methods under the same simulation conditions.

$$\boldsymbol{V_A} = 12 - [12\cos(8660.25t) + 6.39\sin(8660.25t)\,]e^{-5000t}$$

Result and Discussion

Figure 1 illustrates the voltage-time response of a series RLC circuit under underdamped conditions. The oscillatory response with decreasing amplitude reflects the underdamped nature of the system. Both RK4 and RK5 captured this transient behavior; however, RK5 tracked the benchmark solution more closely. This is due to its inclusion of higher-order expansion terms, which reduce local truncation errors. In contrast, RK4, while accurate, accumulates error more rapidly over time.



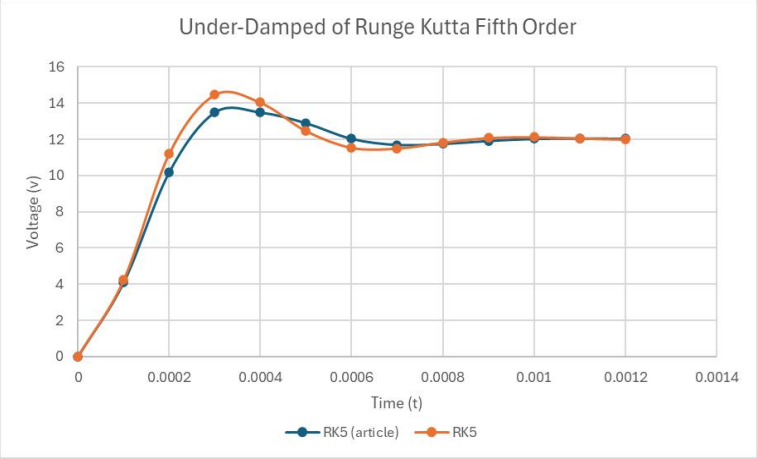


Figure 1: Voltage-Time Response of the Underdamped RLC Series Circuit

Figure 2 compares the voltage response obtained from Euler, RK3, RK4, and RK5 against the benchmark BRK5 solution. The results show that Euler and RK3 deviate significantly from the reference, especially as time progresses. RK5 remains consistently closer to the BRK5 curve than RK4, owing to its higher-order Taylor expansion terms. These additional terms reduce truncation errors and improve fidelity in capturing fast-changing signals, explaining why RK5 outperforms RK4, particularly in oscillatory systems. The x-axis represents simulation time (seconds), while the y-axis represents circuit voltage (volts). This allows direct comparison of how accurately each method tracks the reference response over time.

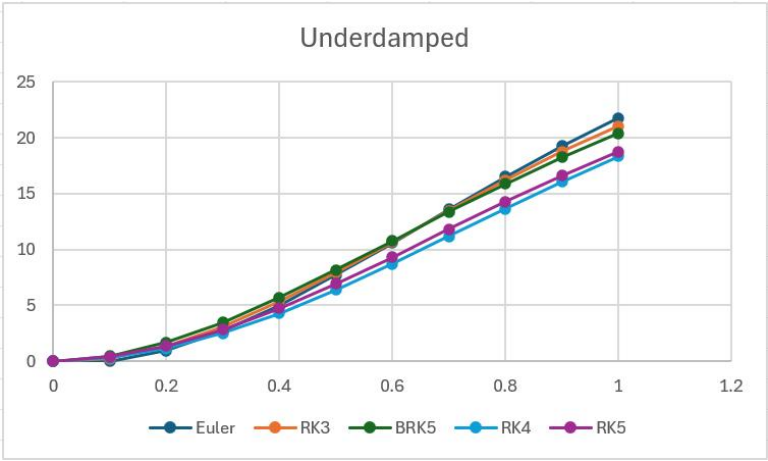


Figure 2: Voltage-Time Response of the Underdamped RLC Parallel Circuit

Table 1 presents the absolute error analysis for the underdamped series RLC circuit over a short simulation window (t = 0.0000 to 0.0012 seconds). The errors are generally small for both RK4 and RK5. Interestingly, RK4 sometimes yields smaller errors than RK5 within this short interval. This indicates that RK4 can perform well in short-duration simulations where computational efficiency is a priority, since its simpler formulation requires fewer function evaluations per step.

Table 1: Error Analysis of RK4 and RK5 of RLC Series Circuit
(Case I: Underdamped)

t	Exact Solution	Abs Error	
		RK4	RK5
0.0000	0	0	0
0.0001	4.0828	0.0828	0.1447
0.0002	10.1925	0.0050	1.0215
0.0003	13.4920	0.0869	0.9744
0.0004	13.8375	0.09532	0.1951
0.0005	12.8952	0.0401	0.4289
0.0006	12.0276	0.0153	0.4899
0.0007	11.6923	0.0350	0.1972
0.0008	11.7481	0.0240	0.0778
0.0009	11.9152	0.00452	0.1588
0.0010	12.0260	0.007187	0.0934
0.0011	12.0517	0.0082915	0.0033
0.0012	12.0310	0.00387	0.0389

Table 2 compares the absolute errors for the underdamped parallel RLC circuit over a longer simulation interval (t = 0.0 to 1.0 seconds). Here, RK5 demonstrates superior performance, with consistently lower errors than RK4 throughout the simulation. The advantage of RK5 becomes more pronounced as time increases, since its higher-order accuracy helps reduce

cumulative truncation error. This makes RK5 more reliable for long-duration simulations or systems where sustained precision is necessary.

Table 2: Absolute Error Analysis of RK4 and RK5 of RLC Parallel Circuit (Case II: Underdamped)

t	Exact Solution	Abs Error	
		RK4	RK5
0.0	0	0	0
0.1	0.73755	0.44349	0.34137
0.2	2.18195	1.03632	0.81197
0.3	4.15289	1.65848	1.30695
0.4	6.48369	2.21999	1.75112
0.5	9.02491	2.66072	2.09789
0.6	11.64659	2.94854	2.32655
0.7	14.23940	3.07625	2.43851
0.8	16.71490	3.05767	2.45295
0.9	19.00490	2.92302	2.40189
1.0	21.06029	2.71405	2.32534

The results also highlight the role of damping in numerical accuracy. In underdamped circuits, oscillations amplify the effect of numerical errors, making higher-order methods like RK5 particularly beneficial. In overdamped or critically damped cases, where responses are smoother, RK4 may provide sufficiently accurate results with lower computational cost. This demonstrates that damping characteristics directly influence the degree to which higher-order accuracy improves numerical performance. For engineers, this means RK4 is well-suited for real-time applications or embedded systems, while RK5 should be preferred in high-precision modeling tasks such as nonlinear circuits, control systems, or long-duration simulations.

Conclusion

This study evaluated the application of the Fourth-Order Runge–Kutta (RK4) and Fifth-Order Runge–Kutta (RK5) methods in modeling the dynamic behavior of RLC electrical circuits. By simulating transient and steady-state responses under underdamped, critically damped, and

overdamped conditions, both methods were shown to be effective for circuit analysis. RK4 offered a balance between simplicity and accuracy, making it particularly suitable for short-duration or real-time simulations where computational efficiency is important. In contrast, RK5 consistently provided higher accuracy, especially in long-duration simulations and under oscillatory conditions where error accumulation is significant.

However, this work has several limitations. The analysis was restricted to linear RLC circuits, and simulations were conducted using Mathematica without adaptive step-size control. Future research should extend this comparison to include nonlinear and stiff circuits, investigate the performance of adaptive Runge–Kutta methods, and benchmark the results using other simulation environments

In conclusion, the findings of this study highlight the practical trade-off between computational efficiency and accuracy. RK4 remains a reliable option for fast, resource-limited applications, while RK5 is preferred when long-term precision and error minimization are critical. This guidance can help engineers and researchers make informed decisions when selecting numerical methods for circuit modeling and simulation.

Acknowledgments

The authors would like to thank the supervisors and faculty members for their continuous support and guidance throughout this research. Special thanks also go to those who provided valuable feedback and encouragement during the preparation of this project.

References

[1] Chen, K., Xiao, B., Wang, C. Y., Liu, X. L., Zhang, X., & Xia, C. (2024). FRKVF: High-accuracy motion interpolation for polishing operations using fourth-order Runge-Kutta and velocity flexibility planning. *Precision Engineering*, *90*, 122–140. https://doi.org/10.1016/j.precisioneng.2024.08.008

[2] Ashgi, R., Andhika, M., Pratama, A., & Purwani, S. (2021). Comparison of Numerical Simulation of Epidemiological Model between Euler Method with 4th Order Runge Kutta Method. *International Journal of Global Operations Research*, *2*(1), 37–44. http://www.iorajournal.org/index.php/ijgor/index

[3] Ezhilarasi, A. (2023). A Study of Runge-Kutta Method for n-th Order Differential Equations. *In Proceeding of 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, (ICAECA 2023)*. https://doi.org/10.1109/ICAECA56562.2023.10200469

[4] Kafle, J., Thakur, B. K., & Bhandari, I. B. (2021). Application of Numerical Methods for the Analysis of Damped Parallel RLC Circuit. *Journal of Institute of Science and Technology*, *26*(1), 28–34. https://doi.org/10.3126/jist.v26i1.37814

[5] Shaikh, A. G., Keerio, U., Shaikh, W., & Sheikh, A. H. (2022). Numerical higher-order Runge-Kutta methods in transient and damping analysis. *International Journal of Advanced and Applied Sciences*, *9*(10), 174–179. https://doi.org/10.21833/ijaas.2022.10.020

[6] Ying, P., Tang, H., Chen, L., Ren, Y., & Kumar, A. (2023). Dynamic modeling and vibration characteristics of multibody system in axial piston pump. In *Alexandria Engineering Journal* (Vol. 62, pp. 523–540). Elsevier B.V. https://doi.org/10.1016/j.aej.2022.07.046

[7] Kazeem Iyanda, F., Tunde, T. A., & Isa, U. (2021). Numerical Comparison Of Runge Kutta (RK5) And New Iterative Method (NIM) For Solving Metastatic Cancer Model. *Malaysian Journal of Computing*, *6*(1), 758–771. https://doi.org/10.24191/mjoc.v6i1.174

OPTIMISING COURIER DELIVERY ROUTES AND TIME TRAVEL USING DIJKSTRA'S ALGORITHM TO FIND THE SHORTEST PATH

Noor Hidayah Mohd Zaki 1[1,*], Nurin Afiqah Bakhtiar 1[1], Nurul Syahida Zanurin 2[2]

[1,*,1,2]Faculty of Computer and Mathematical Science, Universiti Teknologi MARA, Cawangan Seremban, Negeri Sembilan, Malaysia
*email:hidayahzaki@uitm.edu.my

Abstract

This project seeks to find the shortest path and minimum travel time courier delivery by using Dijkstra's Algorithm in Lukut, Port Dickson. Efficient route planning has become more relevant because of the rising demand for fast delivery services due to the growth of e-commerce. This research aims to enhance delivery performance by identifying the shortest path between multiple delivery locations using graph theory. The records of the distance and travel time to nine delivery points were obtained using Google Maps. Nodes are the assumed location of these places, and the edges between locations are the roads with weight in the form of time and distance. Dijkstra's Algorithm was applied in Microsoft Excel to determine the fastest route between Parcelhub, Lukut and the Tasik Villa International Resort Port Dickson. It was found out that the application of this algorithm would assist in reducing delivery time and making the route efficient. The technique cuts down on the amount of fuel used, delays in delivery and the cost of operation. Its outcomes indicate that although only a small area is studied and a set of static data is used, it shows strong potential for practical use in real-life courier services. This project proves that mathematical tools like Dijkstra's Algorithm can help solve real-world logistics problems by providing faster and more reliable delivery routes, improving service quality and customer satisfaction.

Keywords: Dijkstra's Algorithm, shortest path, courier delivery, route optimisations, graph theory

1. Introduction

In recent years, the courier delivery industry has experienced a significant transformation by adopting modern technology to enhance service efficiency. The surge in e-commerce has driven the need for faster, more reliable delivery systems. Companies such as Pos Laju, GD Express, Ninja Van, and J&T Express have responded by optimising their logistics strategies, particularly through route optimisation. Efficient delivery route planning minimises operational costs and improves customer satisfaction. This study focuses on implementing Dijkstra's Algorithm to optimise courier delivery routes in Lukut, Port Dickson, a key area affected by delivery congestion. Dijkstra's Algorithm is a reliable shortest path method in graph theory, offering accurate route decisions by analysing nodes, edges, and associated weights such as time and distance. The use of this algorithm in courier delivery services helps address delivery delays, fuel consumption, and overall logistics cost, thus demonstrating its value in real-world applications.

2. Literature Review

Route optimisation has been approached using a variety of algorithms. Genetic Algorithms (GAs) are commonly used for their ability to explore complex search spaces efficiently, particularly when hybridised with other metaheuristics (Katoch et al., 2021). Ant Colony Optimisation (ACO), inspired by the foraging behaviour of ants, has effectively solved vehicle routing problems and search and rescue scenarios (Morin et al., 2023). The A* Algorithm, combining Dijkstra's method with heuristic functions, provides efficient pathfinding in dynamic environments (Suryadibrata et al., 2019). Bellman-Ford Algorithm supports negative weights and fuzzy logic in decision-making for uncertain networks (Parimala et al., 2021). However, among all of them, Dijkstra's Algorithm stands out for its accuracy and performance in static and non-negative weighted graphs, making it ideal for fixed delivery networks (AlShaikh et al., 2019). Several studies confirm its superiority in route planning, with improvements focusing on multi-objective optimisation, energy efficiency, and dynamic rerouting.

3. Methodology

This study used a three-phase methodology to evaluate and determine the shortest route with minimal travel time in Lukut, Port Dickson. Figure 1 presents the structured process used in this research.

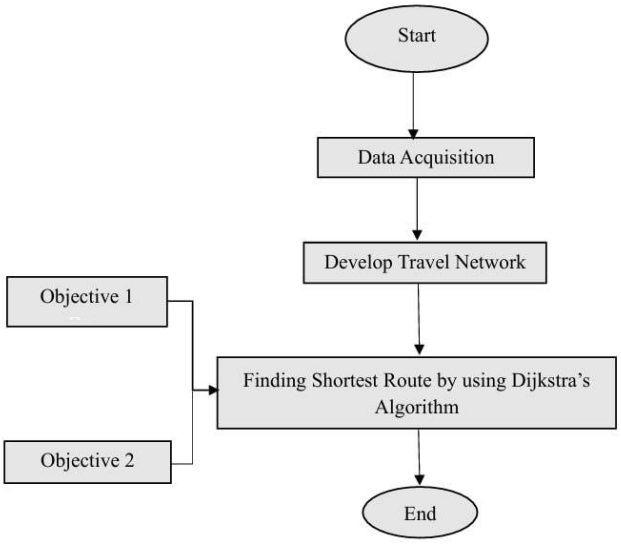


Figure 1: The structured process used in this study

The first phase involved data acquisition from Google Maps, which provided real-time data including distance and time between delivery locations. In Figure 2, the selected delivery points included Parcelhub Lukut, Tasik Villa International Resort, and seven other key sites within a 1.5 km radius of Parcelhub Lukut.



Figure 2: Nine places in Lukut for the delivery van to stop. (source: Google Maps)

In Phase Two, Figure 3 and Figure 4 show the travel network developed using nodes (delivery points) and edges (routes), with each edge assigned a weight representing time or distance.

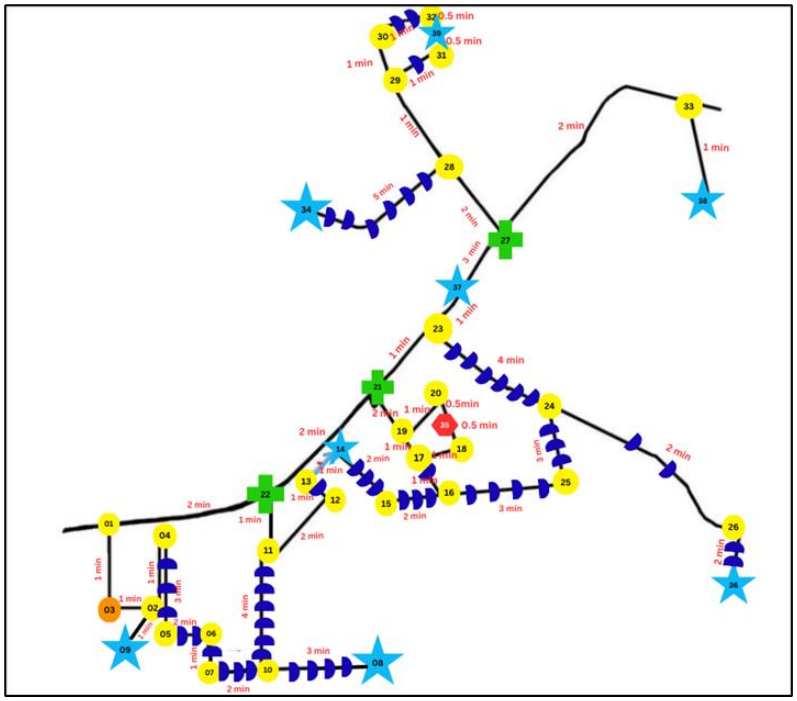


Figure 3: The complete sample travel network with time travel in minutes

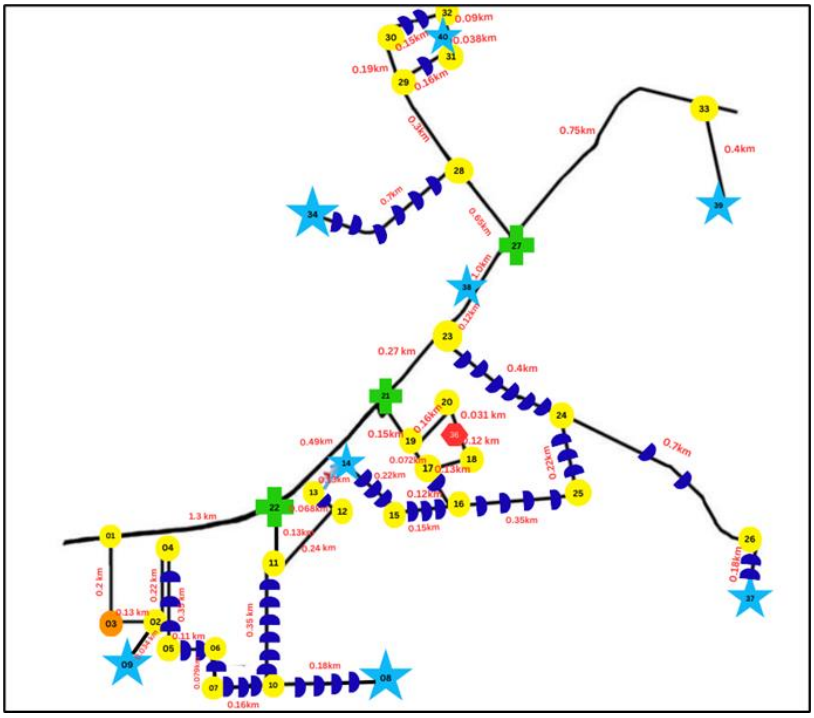


Figure 4: The complete sample travel network with distance travel in kilometres

Phase Three involved applying Dijkstra's Algorithm through Manual Calculation and Microsoft Excel, which allowed visualisation of travel networks and shortest path computation. The algorithm was manually applied in a preliminary study and later validated using Excel Solver for practical implementation. Time penalties were included for traffic lights, junctions, and road conditions to improve the realism of travel time estimation.

4. Equation

Seeta (2021) said that Dijkstra's algorithm works through a series of steps where a labelling process is carried out at each iteration. During this process, each node V is assigned a label representing the shortest path length from the starting node s to V, considering only nodes already included in a designated set. At each iteration, the algorithm selects the node with the smallest label that is not in the set and adds it to the set. This process ensures that the shortest path is obtained. Each node is assigned with Permanent Label (PL), which represents the finalised shortest path length or Temporary Label (TL), which is the neighbour of the currently selected node.

The equation to find the shortest time is:

$$(s, j) = \min(L'_j, L_j + L_{kj}) \quad (1)$$

where,

s = Initial node

L'_j = The value of time travel from the initial node to node j

L_j = Total current minimum time travel

L_{kj} = The time travel from node k to node j

5. Results and Discussion

Table 1 presents the summarised results of the whole study. The "Places" column shows the sequence of delivery points, such as starting from Parcelhub Lukut and ending at Tasik Villa International Resort Port Dickson. The "Route" column shows the specific sequence of nodes, which are identified as the shortest path by the Dijkstra Algorithm. The "Minimum Distance" and "Minimum Time" columns represent the total length and travel duration for each path, respectively. For example, the route from Parcelhub Lukut to Pejabat Pendidikan Daerah Port Dickson has a minimum distance of 2.881 km and requires 17.5 minutes of travel time by using the shortest path, which is the node sequence 35-20-19-21-23-37-27-33-38. This process is consistently applied to each delivery point, ensuring a continuous and optimised route across all delivery locations.

The cumulative result, as presented in the final row of the table, indicates that the total optimised distance for the complete delivery operation is 14.048 km, while the total estimated travel time is 109 minutes. These results show the effectiveness of Dijkstra's Algorithm in significantly reducing delivery route length and time, thus fulfilling the objectives of this study, which are to find the travel network, minimum time and minimum distance across all delivery locations to optimise courier delivery operations.

Table 1: The summarised results of the whole study

Places	Route	Minimum distance (km)	Minimum time (min)
Parcelhub Lukut to Pejabat Pendidikan Daerah Port Dickson.	35-20-19-21-23-37-27-33-38	2.881	17.5
Pejabat Pendidikan Daerah Port Dickson to Atlas Edible Ice Sdn Bhd.	38-33-27-28-29-31-39	2.298	11.5
Atlas Edible Ice Sdn Bhd to Taman Wawasan.	39-31-29-28-34	1.198	9.5
Taman Wawasan to Lukut Police Station.	34-28-27-37	2.35	12.5
Lukut Police Station to AJL Trading.	37-23-24-26-36	1.4	11
AJL Trading to CIMB Bank Lukut	36-26-24-25-16-15-14	1.82	17

CIMB Bank Lukut to SJK(C) Chung Hua.	14-13-12-11-10-08	0.968	13.5
SJK(C) Chng Hua to Tasik Villa International Resort Port Dickson.	08-10-07-06-05-04-02-09	1.133	16.5
TOTAL		14.048	109

6. Conclusion and Recommendations

This study concludes that Dijkstra's Algorithm is an effective solution for optimising courier delivery routes and minimising travel time. The implementation in Lukut, Port Dickson, demonstrates how mathematical algorithms can provide actionable insights into logistics operations. Future research should consider real-time traffic integration and dynamic route adjustments to enhance the system's adaptability. Additionally, expanding the study to cover larger geographic areas and using real-time delivery data can further validate the algorithm's scalability and effectiveness.

References

AlShaikh, M., Laouamer, L., Laouid, A., Bounceur, A., & Hammoudeh, M. (2019). Robust and Imperceptible Medical Image Watermarking Based on Dijkstra Algorithm. ACM International Conference Proceeding Series. https://doi.org/10.1145/3341325.3341991

Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A Review on Genetic Algorithm: Past, Present, and Future. Multimedia Tools and Applications, 80(5), 8091–8126. https://doi.org/10.1007/s11042-020-10139-6

Morin, M., Abi-Zeid, I., & Quimper, C.-G. (2023). Ant Colony Optimization for Path Planning in Search and Rescue Operations. European Journal of Operational Research, 305(1), 53–63. https://doi.org/10.1016/j.ejor.2022.06.019

Parimala, M., Broumi, S., Prakash, K., & Topal, S. (2021). Bellman–Ford Algorithm for Solving Shortest Path Problem of A Network Under Picture Fuzzy Environment. Complex and Intelligent Systems, 7(5), 2373–2381. https://doi.org/10.1007/s40747-021-00430-w

Seeta, V. (2021). Finding the Shortest Path using Dijkstra's Algorithm. International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences, 9(5), 1–1. https://doi.org/10.37082/ijirmps.2021.v09si05.017

Suryadibrata, A., Young, J., & Luhulima, R. (2019). Review of Various A* Pathfinding Implementations in Game Autonomous Agent. 43 IJNMT, VI(1). https://doi.org/10.31937/ijnmt.v6i1.1075

SOLVING HEAT EQUATION USING FINITE DIFFERENCE METHOD

Nur Zeti Azreen Nor Azam[1], NurHayati Mohd Nazeri[1], Azhar Ahmad[2,*] and Norshuhada Samsudin[2]

[1]Fakulti Sains Komputer & Matematik (FSKM),
Universiti Teknologi MARA Cawangan Negeri Sembilan, Malaysia
[2]Jabatan Sains Komputer & Matematik (JSKM),
Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia
*azhar251@uitm.edu.my

Abstract

This paper presents a numerical solution of the one-dimensional heat equation using the Crank-Nicolson finite difference method. The heat equation is a second-order partial differential equation that models the distribution of heat over time in a given medium. Due to the limitations of analytical methods in solving problems with complex boundary conditions, numerical approaches are essential. In this study, we focus on a test problem involving Dirichlet boundary conditions at both ends of the domain. The continuous heat equation is discretized in both space and time to form a system of linear algebraic equations. The Crank-Nicolson scheme is selected for its stability and accuracy. Numerical implementation is carried out using Wolfram Mathematica, which also enables the generation of 3D surface plots for better visualization. Error analysis is performed by comparing the numerical solution to the exact analytical solution using the L_2 norm and maximum absolute error. Results indicate that the method yields accurate and stable approximations for the heat distribution over time.

Keywords: One-dimensional, heat equation, Crank-Nicolson scheme, Dirichlet boundary condition, numerical method

1. Introduction

The heat equation is one of the most widely studied partial differential equations in applied mathematics and physics due to its role in modeling the diffusion of heat over time. It has many real-world applications such as in thermal conduction, fluid dynamics, and material sciences (Mebrate, 2015). The analytical solutions for the heat equation are often limited to simple geometries and boundary conditions, which may not be suitable for more complex or practical situations (Nolasco et al., 2019).

To address these limitations, numerical methods such as the Finite Difference Method (FDM) are commonly used. Among various FDM techniques, the Crank-Nicolson method is known for being unconditionally stable and second-order accurate in both time and space (Mohebbi & Dehghan, 2010). This makes it particularly effective for solving parabolic equations like the heat equation. In this study, the Crank-Nicolson method is used to simulate a one-dimensional heat conduction problem with Dirichlet boundary conditions. The implementation and visualization are carried out using Wolfram Mathematica. The numerical results are then validated through error analysis by comparing them with the known analytical solution.

In this study, the Crank-Nicolson scheme is applied to solve a one-dimensional heat conduction problem with Dirichlet boundary conditions imposed at both ends of the domain. The numerical experiment is carried out using Wolfram Mathematica, which provides a flexible environment for both computation and visualization. The heat distribution is examined through surface plots and the accuracy of the method is evaluated using error metrics such as the L_2 norm and $L\infty$ norm. The goal is to demonstrate the effectiveness of the Crank-Nicolson approach in producing accurate numerical results for heat conduction problems.

2. Literature review

The heat equation is a fundamental partial differential equation (PDE) that models the distribution of temperature over time. Due to the complexity of real-world problems, such as irregular geometries and time-varying boundary conditions, analytical solutions are often impractical, making numerical methods essential (Mebrate, 2015). One of the most widely used techniques is the Finite Difference Method (FDM), which transforms the continuous heat equation into a system of algebraic equations through spatial and temporal discretization (Brezzi et al., 2005; Adamiec-Wojcik et al., 2013). Among numerical schemes, the Crank-Nicolson method, stands out for its second-order accuracy and unconditional stability in both space and time (Liu & Hao, 2022; Mohebbi & Dehghan, 2010). It offers a balanced approach between the explicit and implicit methods, making it especially effective for transient heat conduction. Studies have shown its advantages over Forward Time Centered Space (FTCS) schemes, which suffer from stability limitations (Mojumder et al., 2023; Loskor & Sarkar, 2022). Although Crank-Nicolson requires solving a system of equations at each time step, its accuracy and robustness justify the computational effort.

Boundary conditions also significantly impact simulation accuracy. Dirichlet conditions, which fix temperature at boundaries and Neumann conditions, which fix heat flux are both commonly applied in FDM based heat simulations (Hajrulla et al., 2023; Nolasco et al., 2019). Chai et al. (2020) successfully used finite difference techniques with level set and ghost fluid methos to handle complex domains with mixed boundary conditions, proving the adaptability of FDM in irregular geometries. Furthermore, the distinction between homogeneous and non-homogeneous heat equations is vital. Homogeneous equations, with uniform properties and no internal heat source, are easier to solve and often used in steady-state simulations (Ahmad & Gunel, 2023). In contrast, non-homogeneous equations involve variable material properties or internal heat sources and require more advanced discretization methods. Safari (2024) demonstrated the effectiveness of combining Crank-Nicolson with meshless techniques to handle irregular material distributions and boundary complexities.

In summary, the FDM particularly the Crank-Nicolson scheme remains one of the most reliable, accurate and flexible methods for solving both homogeneous and non-homogeneous heat equations under a variety of boundary conditions.

3. Methodology

The process of solving that one-dimensional heat equation in this study begins with the formulation of the standard heat equation:

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \tag{1}$$

where $T =$ temperature, $t =$ time, $x =$ spatial coordinates, and $\alpha =$ thermal diffusivity are constant.

In real-world applications, exact analytical solutions are often only available for very simple initial and boundary conditions. When dealing with more complex or practical cases, such as irregular domains or mixed boundary types, numerical methods like the Finite Difference Method (FDM) become essential. These methods allow the approximation of solutions by discretizing the continuous domain, enabling simulation and analysis that would otherwise be impossible.

To solve this, the Finite Difference Method (FDM) is used to discretize the spatial and temporal domains. The Crank-Nicolson method is applied as the chosen scheme for time discretization. It is known for its second-order accuracy in both time and space as well as its numerical stability. The equation is thus transformed into a system of algebraic equations that can be solved computationally.

The implementation is carried out in a one-dimensional spatial domain with uniform grid spacing. Wolfram Mathematica is utilized to perform numerical computations and to generate visual outputs, including line plots and surface plots to illustrate the evolution of temperature over time. The accuracy of the results is evaluated through error analysis by comparing the numerical solution with the exact analytical solution using both the L_2 norm and the maximum absolute error, L∞ norm.

To evaluate the accuracy of the numerical solution, two error metrics were used:

- The L2 norm, which measures the root mean square error

$$L_2\ norm = \sqrt{\sum_{i=1}^{N_X} (T_{numerical}(i,t) - T_{analytical}(i,t))^2} \tag{2}$$

- The maximum absolute error (L∞ norm) which identifies the largest deviation between the numerical and analytical solutions

$$L_\infty\ norm = max_i \left| T_{numerical}(i,t) - T_{analytical}(i,t)) \right| \tag{3}$$

4. Results and Discussion

This study examines the temperature distribution along one-dimensional rod governed by the heat equation:

$$\frac{\partial \phi}{\partial t} = \alpha \frac{\partial^2 \phi}{dx^2} \qquad \text{for } 0 \le x \le L,\ t \ge 0 \tag{4}$$

Initial condition $$\phi(x,0) = f(x) \tag{5}$$

Boundary condition $$\phi(0,t) = \phi_0 \tag{6}$$

$$\phi(L,t) = \phi_L \tag{7}$$

The numerical simulation was carried out using Wolfram Mathematica. The spatial domain is divided uniformly from $x = 0$ to $x = 10$, and the Crank-Nicolson scheme is applied at each time step to compute the temperature distribution. Both ends of the rod are held at fixed temperatures (Dirichlet boundary conditions). Figure 1 displays a combined 3D surface plot of the temperature distribution obtained using the Crank-Nicolson method and the exact analytical solution. The visual comparison indicates that both surfaces are nearly identical, highlighting the accuracy of the numerical approach.

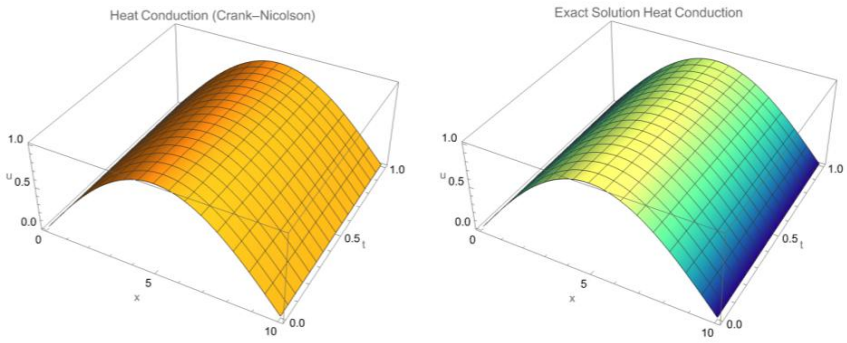


Figure 1: 3D-plot of Crank-Nicolson Solution and the Exact Solution

To evaluate the accuracy of the numerical method, results at selected spatial points and a fixed time level were compared to the exact values. Table 1 presents the exact solution, numerical results and the corresponding absolute errors.

Table 1: Comparison of Crank-Nicolson and Exact Solutions with Absolute Error

x	Exact Solution	FDM (Ucn)	Absolute Error
0	0.0000000	0	0.0000000
1	0.2799750	0.2799755	5.6760409×10^{-7}
2	0.5325441	0.5325451	1.0796471×10^{-6}
3	0.7329840	0.7329855	1.4860068×10^{-6}
4	0.8616744	0.8616761	1.7469058×10^{-6}
5	0.9060181	0.9060199	1.8368054×10^{-6}
6	0.8616744	0.8616761	1.7469058×10^{-6}
7	0.7329840	0.7329855	1.4860068×10^{-6}
8	0.5325441	0.5325451	1.0796471×10^{-6}
9	0.2799750	0.2799755	5.6760409×10^{-7}
10	0.0000000	0	0.0000000
$L_2 - norm$	–	–	0.0000184
$L\infty - norm$	–	–	1.8368054×10^{-6}

The computed error metrics are:

- L_2 norm: 0.0000184
- L∞ norm: 1.8368054×10^{-6}

These small errors further confirm that the numerical solution achieved in this project is both accurate and stable. One reason for the small error in Problem 2 could be due to its Dirichlet boundary conditions, which generally provide better numerical stability by controlling the boundary temperature effectively. Further, discretization parameters including spatial step also contributed significantly to the minimization of numerical errors.

To summarize, the outcome Problem 2 using the graphical comparison, data table analysis, and error norm calculation indicate that the Crank-Nicolson finite difference finds a suitable and reasonable numerical scheme to solve one dimensional heat conduction problem with the Dirichlet boundary conditions.

5. Conclusion

This study demonstrated how the Finite Difference Method (FDM), specifically the Crank-Nicolson scheme, is highly accurate and stable approach for solving the one-dimensional heat equation under Dirichlet boundary conditions. Using Wolfram Mathematica for implementation and visualization, the numerical solution was compared against the exact analytical solution. The results showed strong agreement, supported by low L_2 norm (0.0000184) and $L\infty$ error norm (1.8368054×10^{-6}) highlighting the method's numerical stability and second order accuracy in both time and space. Overall, the Crank-Nicolson method proves to be a robust and reliable approach for modeling heat conduction problems with fixed boundary temperatures. For future work, this approach can be extended to more complex scenarios such as time-varying or Robin boundary conditions, non-uniform grids and higher dimensional problems.

Acknowledgment

The authors would like to express their sincere gratitude to the Fakulti Sains Komputer & Matematik, Universiti Teknologi MARA, for the guidance and support throughout this project. Special thanks are also extended to our supervisors and lecturers for their valuable advice and encouragement.

References

Adamiec-Wójcik, I., Nowak, A., & Wojciech, S. (2013). Comparison of methods for modelling vibrations of collecting electrodes in dry electrostatic precipitators. *Archive of Mechanical Engineering*, *60*(3), 431–449. https://doi.org/10.2478/meceng-2013-0027

Ahmad, M. J., & Gunel, K. (2023). Exploring the Learning Potential of ELM from Finite Difference Solutions for Heat Equation. *IEEE Explore*. https://doi.org/10.1109/pci60110.2023.10325983

Brezzi, F., Lipnikov, K., & Simoncini, V. (2005). A FAMILY OF MIMETIC FINITE DIFFERENCE METHODS ON POLYGONAL AND POLYHEDRAL MESHES. *Mathematical Models and Methods in Applied Sciences*, *15*(10), 1533–1551. https://doi.org/10.1142/s0218202505000832

Chai, M., Luo, K., Wang, H., Zheng, S., & Fan, J. (2020). Imposing mixed Dirichlet-Neumann-Robin boundary conditions on irregular domains in a level set/ghost fluid based finite difference framework. *Computers & Fluids*, *214*, 104772. https://doi.org/10.1016/j.compfluid.2020.104772

Hajrulla, S., Sheko, A., Isufi, H., Bezati, L., & V Lino. (2023). Finite Difference Method and Numerical Simulations on Heat Transfer Problem. In *2nd International Conference on Scientific and Innovative Studies*. All Sciences Academy. https://as-proceeding.com/index.php/icsis

Liu, J., & Hao, Y. (2022). Crank–Nicolson method for solving uncertain heat equation. *Soft Computing*, *26*(3), 937–945. https://doi.org/10.1007/s00500-021-06565-9

Loskor, W. Z., & Sarkar, R. (2022). A Numerical Solution of Heat Equation for Several Thermal Diffusivity Using Finite Difference Scheme with Stability Conditions. *Journal of Applied Mathematics and Physics*, *10*(02), 449–465. https://doi.org/10.4236/jamp.2022.102034

Mebrate, B. (2015). Numerical Solution of a One-Dimensional Heat Equation with Dirichlet Boundary Conditions. *American Journal of Applied Mathematics*, *3*(6), 305. https://doi.org/10.11648/j.ajam.20150306.20

Mohebbi, A., & Dehghan, M. (2010). High-order compact solution of the one-dimensional heat and advection–diffusion equations. *Applied Mathematical Modelling*, *34*(10), 3071–3084. https://doi.org/10.1016/j.apm.2010.01.013

Mojumder, M. S. H., Haque, M. N., & Alam, M. J. (2023). Efficient Finite Difference Methods for the Numerical Analysis of One-Dimensional Heat Equation. *Journal of Applied Mathematics and Physics*, *11*(10), 3099–3123. https://doi.org/10.4236/jamp.2023.1110204

Nolasco, C., Jácome, N. J., & Hurtado-Lugo, N. A. (2019). Solution by numerical methods of the heat equation in engineering applications. A case of study: Cooling without the use of electricity. *Journal of Physics Conference Series*, *1388*(1), 012034. https://doi.org/10.1088/1742-6596/1388/1/012034

Safari, F. (2024). A stable numerical scheme for solving heat transport equations on the microscopic and cracked domains. *International Communications in Heat and Mass Transfer*, *152*, 107315. https://doi.org/10.1016/j.icheatmasstransfer.2024.107315

Multi-Choice Goal Programming Approach for Multi-Objective Portfolio Optimization Problem

Aisyah Safiyah Abdul Halim[1], Muhammad Danial Iskandar Nizam[1], Mazura Mokhtar[2,*]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan.*
[2]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Cawangan Pahang, Kampus Raub, Pahang.*
[*]mazura_mokhtar@uitm.edu.my

Abstract

In real-world investment scenarios, investors tend to make decisions with regard to all the criteria they want including return, risk, and liquidity all together in simple terms to enhance portfolio optimization. In more complex decision-making contexts, they may also have multiple aspirations or preferences associated with these various goals. Thus, the aim of this study is to apply Multi-Choice Goal Programming (MCGP) approach for solving a multi-objective portfolio optimization (MOPO) problem. The model considers three objectives which are maximizing return, minimizing risk and maximizing liquidity. The model also incorporates environmental, social and governance (ESG) constraints alongside other practical constraints, including cardinality, sector, floor and ceiling constraints to reflect real-world investment conditions. Computational experiments were performed to verify and validate the model, using a dataset of 30 companies listed on Bursa Malaysia, which includes monthly return and turnover rates over the period from January 2020 to December 2023. The data was prepared in Excel before being processed in MATLAB2018a. The model was tested across four scenarios to examine the impact of changing priority weight on the solutions. The results show that the MCGP model could generate optimal portfolios that meet the specified target goals. The results also highlight the effectiveness of the MCGP model in providing investors with the flexibility to define their target goals as interval values.

Keywords*:* Goal Programming, Multi-choice Goal Programming, Portfolio Optimization

Introduction

Portfolio optimization is a fundamental concept in finance that involves choosing the best combination of assets to achieve high returns while managing a specific level of risk. Since high returns are typically associated with high risk, investors often diversify their investments across multiple assets to reduce overall portfolio risk. This optimization process uses mathematical models to determine the most optimal or efficient asset allocation. The benefits include improved portfolio performance, enhanced returns through allocation synergies, and reduced risk through diversification.

The foundation of portfolio optimization was laid by Harry Markowitz in 1952 with the introduction of the Mean-Variance model, marking the beginning of Modern Portfolio Theory (MPT) [1]. This model formulates asset allocation as a quadratic programming problem which

aims to maximize return and minimize risk. However, solving such problems for large portfolios can be extremely difficult. To address this, several researchers such as Sharpe [2] and Speranza [3] proposed linear programming (LP) approximations of the model. Although the LP approximation simplifies computations, these models are limited to a single objective, which only focuses on optimizing the risk-return trade-off, while real-world investment decisions often involve multiple conflicting objectives.

To accommodate more complex investor preferences, Goal Programming (GP) has emerged as an effective approach for handling multi-objective portfolio optimization problems since it allows for the inclusion of multiple goals, offering flexibility. However, the use of fixed target levels in this model can be restrictive when multiple acceptable outcomes exist. Therefore, to overcome this limitation, this study adopts the Multi-Choice Goal Programming (MCGP) model introduced by Chang [4] to solve the multi-objective portfolio optimization problem. This model enhances traditional GP by allowing multiple aspiration levels for each goal, making it more adaptable to real-world portfolio management scenarios that involve diverse and conflicting objectives.

Literature Review

Multi-Choice Goal Programming (MCGP) is a mathematical optimization approach that improves upon traditional Goal Programming (GP). While traditional GP focuses on achieving a single target for each goal, the primary aim of MCGP is to address complex decision-making scenarios in which decision-makers (DMs) may have multiple aspirations or choices associated with various goals. The concept of MCGP was first introduced by Chang [5] and has since been further developed to facilitate decision-making in uncertain environments.

One significant advantage of MCGP is its ability to allow DMs to express their preferences more accurately by assigning multiple aspiration levels to each goal. This flexibility enables the integration of diverse goals into a single optimization framework [6], making it particularly suitable for multi-objective optimization problems where trade-offs are necessary [7]. Moreover, MCGP effectively addresses uncertainties in decision-making environments by providing a structured approach to incorporating uncertainty through multi-choice parameters. For instance, stochastic elements can be integrated into the programming model, allowing organizations to optimize their decisions under varying conditions. This capability is particularly beneficial in fields such as supply chain management and project portfolio optimization [8-9].

MCGP has been widely applied across various sectors to solve complex and multi-objective decision-making problems. In the project portfolio management, Zhai et al. [10] demonstrated the use of MCGP in improving decision-making by analyzing project interactions. In the supply chain and logistic sector, Rosyidi et al. [11] applied MCGP for supplier and carrier selection, incorporating corporate social responsibility (CSR) considerations. Similarly, Jadhav et al. [12] used it to develop a decision support system aimed at reducing risks and improving customer satisfaction.

In agriculture, MCGP has supported crop and resources optimization such as crop planning, livestock management, and environment adaptation. A study by Mohd Nawawi et al. [13] focused on maximizing crop yield under resource constraints, while Silva [9] applied a revised MCGP model for planning sugarcane harvests. In healthcare, Attari et al. [14] used MCGP to

reduce costs and meet patient demands by minimizing blood product waste. Hezam et al. [15] explored sustainable resource allocation in hospitals through a neutrophilic goal programming model.

The education and finance sectors have also benefited from MCGP. For example, Zuhanda et al. [16] optimized educational program costs and profit targets. In finance, Das et al. [17] used MCGP to manage limited investment resources across goals such as retirement and education. Kamil et al. [18] applied MCGP for Islamic banking resource allocation in line with Shariah principles. Alam [19] built a financial planning model for the case of Saudi Basic Industries Corporation (SABIC) to minimize expenses and maximize profit, while Sharma et al. [20] developed a loan allocation model with sensitivity analysis.

In summary, MCGP has proven to be a flexible and effective approach in addressing multi-objective problems across sectors such as healthcare, agriculture, supply chain, education and finance. Its ability to accommodate multiple aspiration levels makes it a valuable tool for realistic and practical decision-making in both public and private sectors.

Methodology

The methodology of this study consisted of three stages. Stage 1 focused on formulating the Multi-Objective Portfolio Optimization (MOPO) model through four steps. First, data collection was carried out using a dataset of 30 stocks traded in Bursa Malaysia. It included the average monthly return and turnover rates from January 2020 to December 2023, as well as Environmental, Social, and Governance (ESG) scores for each stock in 2023, all obtained from the *Eikon Refinitiv* platform. Secondly, the identification of parameters and decision variables was conducted, forming the basis for MOPO model formulation. Third, the model's constraints were established. This study considered five constraints such as ESG, floor and ceiling, class, cardinality, and budget constraints. Lastly, the objective functions of the portfolio were defined, which include maximizing return, minimizing risk, and maximizing liquidity.

Stage 2 involved formulating the MCGP model to solve the MOPO problem. In this stage, the MOPO model obtained from stage 1 was transformed into a single objective problem using the MCGP approach proposed by Chang [4]. This stage consisted of two steps. The first step is to convert the objective function of the model into goal constraints by assigning positive and negative deviation variables to each objective function. Second, the achievement function of the model was constructed by minimizing the unwanted deviation variable for each goal, ensuring the model prioritizes the closest possible attainment of each target goal. According to Chang [4], for '*the more, the better*' target goal, the MCGP-achievement can be formulated as follows:

$$\text{Minimize} \quad \sum_{i=1}^{m} \left[w_i \left(d_i^+ + d_i^- \right) + \alpha_i \left(e_i^+ + e_i^- \right) \right] \tag{1}$$

subject to,

$$f_i(X) - d_i^+ + d_i^- = y_i \qquad i = 1, 2, \ldots, n \tag{2}$$

$$y_i - e_i^+ + e_i^- = g_{i,\max} \qquad i = 1, 2, \ldots, n \qquad (3)$$

$$g_{i,\min} \leq y_i \leq g_{i,\max} \qquad (4)$$

$$d_i^+, d_i^-, e_i^+, e_i^- \geq 0 \qquad i = 1, 2, \ldots, n \qquad (5)$$

$$X \in F$$

For the case of '*the less, the better',* the MCGP-achievement is formulated as

$$\text{Minimize} \quad \sum_{i=1}^{m} \left[w_i \left(d_i^+ + d_i^- \right) + \alpha_i \left(e_i^+ + e_i^- \right) \right] \qquad (6)$$

subject to,

$$f_i(X) - d_i^+ + d_i^- = y_i \qquad i = 1, 2, \ldots, n \qquad (7)$$

$$y_i - e_i^+ + e_i^- = g_{i,\min} \qquad i = 1, 2, \ldots, n \qquad (8)$$

$$g_{i,\min} \leq y_i \leq g_{i,\max} \qquad (9)$$

$$d_i^+, d_i^-, e_i^+, e_i^- \geq 0 \qquad i = 1, 2, \ldots, n \qquad (10)$$

$$X \in F$$

where $g_{i,\min}$ and $g_{i,\max}$ are the upper and lower bounds, respectively for the *i*-th aspiration level y_i, while $f_i(X)$ is the linear function of $x_1, x_2, \ldots, x_n$ for *i*-th goal. d_i^+ and d_i^- are the positive and negative deviations attached to the *i*-th goal, $\left| f_i(X) - y_i \right|$ in Eq. 2 and Eq. 7, respectively. e_i^+ and e_i^- are the positive and negative deviations attached to $\left| y_i - g_{i,\max} \right|$ and $\left| y_i - g_{i,\min} \right|$. w_i and α_i are the weights attached to the sum of the deviations, while F is a feasible set.

The last stage, Stage 3, focused on the validation of the MCGP model, which involved conducting computational experiments and sensitivity analysis. Computational experiments were carried out for imaginary investors using a real dataset and implemented in MATLAB 2018a software. A sensitivity analysis was performed to identify the changes in parameters that affect the results of the model. The model was tested by changing the cardinality parameters or number of selected stocks in the portfolio to observe how the changes influence the model's outcomes and overall portfolio performance.

The Multi-Objective Portfolio Optimization (MOPO) Model

To describe the multi-objective portfolio optimization model conveniently, all the index, parameters, and decision variables used in this study are listed as follows:

Index

i : Index for each stock $(i = 1, 2, \ldots, n)$

Parameters

e_i	:	The ESG score of the i-th stock
ε_i	:	The minimum proportion of budget that can be invested in stock i
δ_i	:	The maximum proportion of budget that can be invested in stock i
L_m	:	The lower proportion limit for class m
U_m	:	The upper proportion limit for class m
C_m	:	Mutually exclusive sets of stocks $(m = 1, 2, \ldots, M)$
r_i	:	The expected rate of return of stock i
r_{it}	:	The actual rate of return determined from the historical data $(t = 1, 2, \ldots, T)$
l_i	:	The liquidity or the turnover rate of stock i
P_t	:	The unbounded variables to represent the realizations of the portfolio return under the scenario t
E	:	The required total ESG score
k	:	The number of stocks held in the portfolio

Decision variables

x_i	The proportion of stock i in the portfolio
z_i	The binary variable which takes value 1 if stock i is selected in the portfolio, 0 otherwise

Thus, the multi-objective portfolio optimization model is formulated as follows:

$$\text{Maximize} \quad R(x) = \sum_{i=1}^{n} r_i x_i \tag{11}$$

$$\text{Minimize} \quad S(x) = \frac{1}{T} \sum_{t=1}^{T} P_t \tag{12}$$

$$\text{Maximize} \quad L(x) = \sum_{i=1}^{n} l_i x_i \tag{13}$$

subject to,

$$\sum_{i=1}^{n} e_i x_i \geq E \tag{14}$$

$$\varepsilon_i z_i \leq x_i \leq \delta_i z_i \qquad i = 1, 2, \ldots, n \tag{15}$$

$$L_m \leq \sum_{i \in C_m} x_i \leq U_m \qquad m = 1, 2, \ldots, M \tag{16}$$

$$\sum_{i=1}^{n} z_i = k \tag{17}$$

$$\sum_{i=1}^{n} x_i = 1 \tag{18}$$

$$P_t \geq -\sum_{i=1}^{n} (r_{it} - r_i)x_i \qquad t = 1, 2, \ldots, T \tag{19}$$

$$P_t \geq 0 \qquad t = 1, 2, \ldots, T \tag{20}$$

$$z_i \in \{0, 1\} \qquad i = 1, 2, \ldots, n \tag{21}$$

$$x_i \geq 0 \qquad i = 1, 2, \ldots, n \tag{22}$$

In the above model, Eq. 11 denotes the objective function that intends to maximize the portfolio return. Eq. 12 minimizes the portfolio risk while Eq. 13 maximizes the portfolio liquidity. Constraint (14) requires limiting investment in companies with poor environmental, social and governance practices. Constraint (15) determines the minimum and maximum bounds of any stock that can be included in a portfolio. Constraint (16) keeps the portfolio well diversified by limiting total number of stocks in a particular class or sector. Constraint (17) is the cardinality constraints which limit the number of stocks in a portfolio. Constraint (18) ensures the total weight of the portfolio remains equal to 1 as limiting the invested money among all stocks. Constraints (19) and (20) are the mean semi-absolute deviation to measure how much the portfolio performs either overperforms or underperforms. Constraint (21) is the binary variable, while constraint (22) is the non-negative constraint.

The Multi-Choice Goal Programming (MCGP) Model

The MOPO model was solved by transforming the model into a single objective problem using the MCGP approach proposed by Chang [4]. The model is as follows:

Minimize $$w_1 d_1^- + w_2 d_2^+ + w_3 d_3^- + e_1^+ + e_1^- + e_2^+ + e_2^- + e_3^+ - e_3^- \tag{23}$$

subject to,

$$\sum_{i=1}^{n} r_i x_i - d_1^+ + d_1^- = y_1 \tag{24}$$

$$y_1 - e_1^+ + e_1^- = g_{1,\max} \text{ for } |y_1 - g_{1,\max}| \tag{25}$$

$$g_{1,\min} \leq y_1 \leq g_{1,\max} \text{ for bound of } y_1 \tag{26}$$

$$\frac{1}{T}\sum_{t=1}^{T} P_t - d_2^+ + d_2^- = y_2 \tag{27}$$

$$y_2 - e_2^+ + e_2^- = g_{2,\min} \text{ for } |y_2 - g_{2,\min}| \tag{28}$$

$$g_{2,\min} \leq y_2 \leq g_{2,\max} \text{ for bound of } y_2 \tag{29}$$

$$\sum_{i=1}^{n} l_i x_i - d_3^+ + d_3^- = y_3 \tag{30}$$

$$y_3 - e_3^+ + e_3^- = g_{3,\max} \text{ for } \left| y_3 - g_{3,\max} \right| \tag{31}$$

$$g_{3,\min} \leq y_3 \leq g_{3,\max} \text{ for bound of } y_3 \tag{32}$$

$$d_1^+, d_1^-, d_2^+, d_2^-, d_3^+, d_3^-, e_1^+, e_1^-, e_2^+, e_2^-, e_3^+, e_3^- \geq 0 \tag{33}$$

Constraints (14) – (22)

In the above model, Eq. 23 denotes the formulation model of the MCGP achievement function. Constraints (24) until (26) are the conversion of portfolio return into goal constraints by adding positive and negative deviations, aspiration level of the goal, and lower and upper bounds. Constraints (27) to (29) formalize the transformation of portfolio risk into goal constraints by incorporating upper and lower bounds relative to the aspiration level, along with corresponding positive and negative deviation variables. Constraints (30) to (32) represent the transformation of the objective function aimed at maximizing portfolio liquidity into a set of goal constraints, while constraint (33) is the non-negative constraint.

Results and Discussion

In the computational experiments, the number of selected stocks was fixed at eight as selecting too few or an excessively large number of stocks is generally not recommended due to its adverse impact on portfolio diversification. The maximum and minimum proportion of budget that can be invested in each selected stock are 30% and 5% respectively. The upper bound and lower bound for consumer sector were set at 40% and 20%, respectively. Meanwhile, the upper bound for energy sector, industrial sector and property products were at 30%, 40% and 30% each. As for the healthcare sector, the minimum proportion of budget that could be invested was 10%. The value for the ESG score parameter is 60% and lastly, the number of periods is 48. The upper bound and lower bound values for each goal are presented in Table 1 below.

Table 1. The value of lower bound and upper bound for each goal

Goal	Parameter	Symbol	Value
Return	Upper bound	$g_{1,max}$	0.05
	Lower bound	$g_{1,min}$	0.01
Risk	Upper bound	$g_{2,max}$	0.03
	Lower bound	$g_{2,min}$	0.003
Liquidity	Upper bound	$g_{3,max}$	0.5
	Lower bound	$g_{3,,min}$	0.1

To explore the impact of changing the weight of the goals on the solution, four runs of the MCGP were performed, each time with different weight as shown in Table 2. Each run produced an optimal portfolio.

Table 2. The weight assigned to return, risk and liquidity goals

Run	Return	Risk	Liquidity
Run 1 (Portfolio 1)	0.5	0.3	0.2
Run 2 (Portfolio 2)	0.7	0.2	0.1
Run 3 (Portfolio 3)	0.3	0.5	0.2
Run 4 (Portfolio 4)	0.3	0.2	0.5

Table 3 presents a comparison of the optimal values obtained for each run. Portfolio 1 produced the best optimal values of 0.0218 for return, 0.0292 for risk, and 0.1238 for liquidity. For Portfolio 2, the best optimal values were 0.0239 for return, 0.0279 for risk, and 0.1125 for liquidity. Portfolio 3 yielded the best optimal values of 0.0209 for return, 0.0274 for risk, and 0.1213 for liquidity. Portfolio 4 resulted in optimal values of 0.0200 for return, 0.0316 for risk, and 0.1285 for liquidity. Portfolio 2 achieved the highest optimal value for return, attributable to the greater weight assigned to this goal. The lowest value for risk goal was obtained in Portfolio 3, as this portfolio assigned a higher weight for risk goal. As for the liquidity, the highest value was achieved in Portfolio 4 since this portfolio assigned the highest weight for liquidity. Thus, it can be concluded that changing priority weight of the goals gives effect on the optimal solutions of the MCGP model.

Table 3: The comparison of optimal value for all portfolios.

Goal	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4
Return	0.0218	0.0239	0.0209	0.0200
Risk	0.0292	0.0279	0.0274	0.0316
Liquidity	0.1238	0.1125	0.1213	0.1285

Table 4 presents the results of sensitivity analysis. Reducing the number of selected stocks, k from 8 to 6 causes a slight decrease in the portfolio return but resulting in a greater portfolio risk and liquidity. This may be due to the limited number of stocks with high returns that can be selected in the portfolio. However, when k increased from 8 to 10, the portfolio returns slightly increased while risk and liquidity are decreased. This indicates that the portfolio with 10 stocks performed better than with 6 and 8 stocks, since it gave the highest return and the lowest risk.

Table 5 shows the optimal value for Portfolio 2 when the number of selected stocks is changing. A similar result as Portfolio 1 was obtained for Portfolio 2. When k decreased from 8 to 6, portfolio return slightly decreased, while risk and liquidity increased. Next, when k increased from 8 to 10, the return increased, and both risk and liquidity decreased. It was also found that, a portfolio with 10 stocks performed better compared to 6 and 8 stocks because it gives high return and the lowest risk.

Table 4: The comparison of optimal value for portfolio 1 in different selected number of stocks

Number of selected stocks (k)	Portfolio 1		
	Return	Risk	Liquidity
6	0.0211	0.0311	0.1352
8	0.0218	0.0292	0.1238
10	0.0222	0.0270	0.1103

Table 5: The comparison of optimal value for portfolio 2 in different selected number of stocks.

Number of Selected stocks (k)	Portfolio 2		
	Return	Risk	Liquidity
6	0.0233	0.0286	0.1237
8	0.0239	0.0279	0.1125
10	0.0235	0.0257	0.1016

Meanwhile, in Table 6, the results pattern for Portfolio 3 is different from Portfolios 1 and 2. When k decreased from 8 to 6, the optimal values for return, risk and liquidity all increased. In contrast, when k increased from 8 to 10, the value for return, risk and liquidity all decreased. Thus, a portfolio with 10 stocks performed better than 6 and 8 because it gives the lowest risk value.

Table 6: The comparison of optimal value for portfolio 3 in different selected number of stocks.

Number of selected stocks (k)	Portfolio 3		
	Return	Risk	Liquidity
6	0.0212	0.0283	0.1285
8	0.0209	0.0274	0.1213
10	0.0208	0.0256	0.1103

Based on Table 7, when k decreased from 8 to 6, the optimal values for all goals increased. However, when k increased from 8 to 10, the portfolio return is increased while risk and liquidity are decreased. Thus, a portfolio which consists of 6 stocks performed better compared to 8 and 10 because the optimal solution gives the highest value of liquidity and return.

Table 7: The comparison of optimal value for portfolio 4 in different selected number of stocks.

Number of selected stocks (k)	Portfolio 4		
	Return	Risk	Liquidity
6	0.0210	0.0323	0.1364
8	0.0120	0.0316	0.1285
10	0.0176	0.0313	0.1202

Conclusion and Recommendation

This study has formulated a model for multi-objective portfolio optimization problem by considering three objectives which are maximizing portfolio's return, minimizing portfolio's risk and maximizing portfolio's liquidity. Several real-life constraints such as cardinality constraint, sector constraint as well as floor and ceiling constraints were also considered, thus adding to the realism of the model. The model was verified and validated using real data. The results from computational experiments have verified that the MCGP models could be used for generating a set of more realistic and flexible optimal solution in solving multi-objective portfolio optimization problem. However, a few recommendations need to be highlighted for future research to get a better result. The first recommendation is to consider a larger number of stocks, which is hundreds or thousands of stocks, instead of 30 stocks. Having a broader number of stocks will improve diversification and provide a more comprehensive representation of real market situations. In addition, including other real-life constraints such as minimum transaction lot instead of proportion reflects the real-life trading situation.

Acknowledgments

The authors would like to express their sincere gratitude to all individuals who contributed to the success of this study, particularly their families and friends, for their unwavering support and encouragement throughout this journey. Special appreciation is also extended to the management of Universiti Teknologi MARA, Seremban Campus, for their invaluable assistance and support in facilitating the collaboration essential to this research.

References

[1] Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(11), 77-91.

[2] Sharpe, W. F. (1971). A linear programming approximation for the general portfolio analysis problem. *Journal of Financial and Quantitative Analysis*, 6(5), 1263-1275.

[3] Speranza, M. G. (1993). Linear programming models for portfolio optimization. *Finance*, 14, 107-123.

[4] Chang, C.-T. (2008). Revised multi-choice goal programming. *Applied Mathematical Modelling*, 32(12), 2587–2595. https://doi.org/10.1016/j.apm.2007.09.008

[5] Chang, C. T. (2007). Multi-choice goal programming. *Omega*, 35(4), 389-396. https://doi.org/10.1016/j.omega.2005.07.009

[6] El-Baz, M. and Mesallam, Y. (2012). TOPSIS and multi-choice goal programming approach supplier selection in supply chain management. *The Egyptian International Journal of Engineering Sciences and Technology*, 15, 154-166.

[7] Patro, K. K., Acharya, M., Biswal, M. P., & Acharya, S. (2015). Computation of a multi-choice goal programming problem. *Applied Mathematics and Computation*, 271, 489-501. https://doi.org/10.1016/j.amc.2015.09.030

[8] Chang, C., Wu, W., & Lin, S. (2021). Fuzzy multi-choice goal programming and artificial bee colony algorithm for triangular and trapezoidal membership functions. *IEEE Access*, 9, 95267-95281. https://doi.org/10.1109/access.2021.3093938

[9] Silva, A. F. da, Marins, F. A. S., & Dias, E. X. (2015). Addressing uncertainty in sugarcane harvest planning through a revised multi-choice goal programming model. *Applied Mathematical Modelling*, 39(18), 5540–5558. https://doi.org/10.1016/j.apm.2015.01.007

[10] Zhai, S.-L., Wu, X.-L., Wang, S.-Y., & Zhao, T. (2021). Application of interaction effect multichoice goal programming in project portfolio analysis. *Mathematical Problems in Engineering*, 2021, 1-10. https://doi.org/10.1155/2021/1863632

[11] Rosyidi, C. N., Khasanah, A. M., & Laksono, P. W. (2022). Goal programming model for joint decision making of order allocation, supplier selection, and carrier selection considering corporate social responsibility. *Jurnal Teknik Industri*, 24(1), 45–52. https://doi.org/10.9744/jti.24.1.45-52

[12] Jadhav, P., Shelke, A., & Sonar, C. D. (2022). Supply chain risk mitigation: Rescheduling the risky suppliers using multi criteria linear goal programming. *AIP Conference Proceedings*, 2424, 050006. https://doi.org/10.1063/5.0076797

[13] Mohd Nawawi, M. K., Chaloob, I. Z., Ahmed, J. S., Ramli, R., & Sufahani, S. F. (2022). Maximization of strategic crops production in Iraq with fuzzy goal programming. *Universal Journal of Agricultural Research*, 10(1), 20–26. https://doi.org/10.13189/ujar.2022.100102

[14] Attari, M. Y. N., Pasandide, S. H. R., Agaie, A., & Niaki, S. T. A. (2017). Presenting a stochastic multi choice goal programming model for reducing wastages and shortages of blood products at hospitals. *Journal of Industrial and Systems Engineering*, 10, 81–96.

[15] Hezam, I. M., Taher, S. A., Foul, A., & Alrasheedi, A. F. (2022). Healthcare's sustainable resource planning using neutrosophic goal programming. *Journal of Healthcare Engineering*, 2022, 1–23. https://doi.org/10.1155/2022/3602792

[16] Zuhanda, M. K., Suwilo, S., Sitompul, O. S., & Mardingsih, M. (2022). Goal programming method in optimizing course student admission, operational costs and profits. *Journal of Informatics and Telecommunication Engineering*, 5(2), 286-294. https://doi.org/10.31289/jite.v5i2.6072

[17] Das, S. R., Ostrov, D., Radhakrishnan, A., & Srivastav, D. (2021). Dynamic optimization for multi-goals wealth management. *Journal of Banking & Finance*, 140, 106192 https://doi.org/10.1016/j.jbankfin.2021.106192

[18] Kamil, K. H., Ismail, A. G., Shahimi, S., & Isa, Z. (2020). Multi-choice goal programming model for optimal financial resources in Islamic bank. *Jurnal Pengurusan*, 59, 25-35. https://doi.org/10.17576/pengurusan-2020-59-04

[19] Alam, T. (2022). Modeling and analyzing a multi-objective financial planning model using goal programming. *Applied System Innovation*, 5(6), 128-137. https://doi.org/10.3390/asi5060128

[20] Sharma, D. K., Ghosh, D., & Alade, J. A. (2002). A goal programming model for the best possible solution to loan allocation problems. *Korean Journal of Computational & Applied Mathematics*, 9(1), 197-211. https://doi.org/10.1007/bf03012349

SOLVING POISSON'S EQUATION USING THE FINITE DIFFERENCE METHOD

Nur Fatin Syuhada Mohd Hasrulnizam[1], Nur Syazwani Mohd Azmee[1] , Azhar Ahmad[2,*] and Sharifah Sarimah Syed Abdullah[2]

[1] College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Malaysia
[2]Department of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia

[*]azhar251@uitm.edu.my

Abstract

This paper discusses the numerical solution of Poisson's and Laplace equations using the Finite Difference Method (FDM). The two equations were solved two-dimensionally in a rectangular domain under Dirichlet boundary conditions. A five-point stencil scheme was used to discretize the equations, and the linear systems obtained were solved using the Mathematica function 'LinearSolve' . The primary objective was to test the method with various grid sizes (n = 5, 10, 15, 20) and compare the precision and reliability of the two problems—one with a source term (Poisson) and one without (Laplace). The findings revealed that the method produced accurate and consistent solutions, with error decreasing as the grid size increased. Surface plots verified the smoothness of the solution on finer grids, and error analysis confirmed the convergence behaviour. This project shows that FDM is a stable and viable technique for solving boundary value problems and is useful to students and researchers working in numerical analysis and partial differential equations.

Keywords: Finite Difference Method, Poisson's Equation, Laplace Equation, Dirichlet Boundary Condition, Numerical Analysis

1. Introduction

Poisson's equation, written as $\nabla^2\phi = f$, is a fundamental partial differential equation (PDE) appearing in fields such as electrostatics, heat conduction, fluid dynamics, and gravitational modelling. It models the interaction between a potential function ϕ and a source term f in a spatial domain. While analytical solutions are possible in simple geometries, numerical methods are required for complex domains.

The Finite Difference Method (FDM) is a widely used numerical technique due to its simplicity and effectiveness, especially on regular grids. It discretizes the domain into grid points and approximates derivatives using finite differences, converting the PDE into a system of algebraic equations. This system can be solved using direct or iterative methods. In this study, a direct solver is employed due to its effectiveness for small to medium-sized problems.

This project focuses on solving Poisson's and Laplace equations using FDM under Dirichlet boundary conditions. By varying the grid sizes, we analyse the accuracy and convergence of the method.

2. Literature Review

The development of FDM dates back to Richardson (1928), who introduced finite difference approximations for solving differential equations numerically. Since then, FDM has become an essential tool for solving PDEs, especially for structured domains (Thomas, 1995).

Fornberg (1998) highlighted limitations of FDM in non-regular domains and suggested adaptive grid methods to enhance accuracy and efficiency. Briggs et al. (2000) discussed multigrid techniques to accelerate convergence. Zaman (2022) explored matrix-based FDM implementations for Poisson's equation, supporting the approach used in this study.

These studies validate the use of five-point stencil schemes with Dirichlet conditions for structured grids, which forms the foundation for the methodology applied here.

3. Methodology

3.1. Governing Equation

This study focuses on solving:

- Poisson's Equation:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -e^{xy}(x^2 + y^2) \quad (1)$$

- Laplace's Equation:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad (2)$$

The exact solution for the Poisson equation is $u(x, y) = e^{-xy}$ which is used to validate the numerical result.

3.2. Finite Difference Discretization

The two-dimensional domain is discretized into an $n \times n$ grid. The five-point stencil approximation for the Laplacian operator is:

$$\frac{u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j}}{h^2} = f_{i,j} \quad (3)$$

where $h = \frac{1}{n+1}$ is the grid spacing.

This leads to a linear system of the form $Au = f$, where A is the coefficient matrix, u is the unknown potential vector, and f is the source vector. The system is solved using Mathematica's 'LinearSolve' .

3.3. Numerical Parameters and Error Analysis

Grid sizes tested were n = 5, 10, 15, and 20. For each case, the absolute error was computed as:

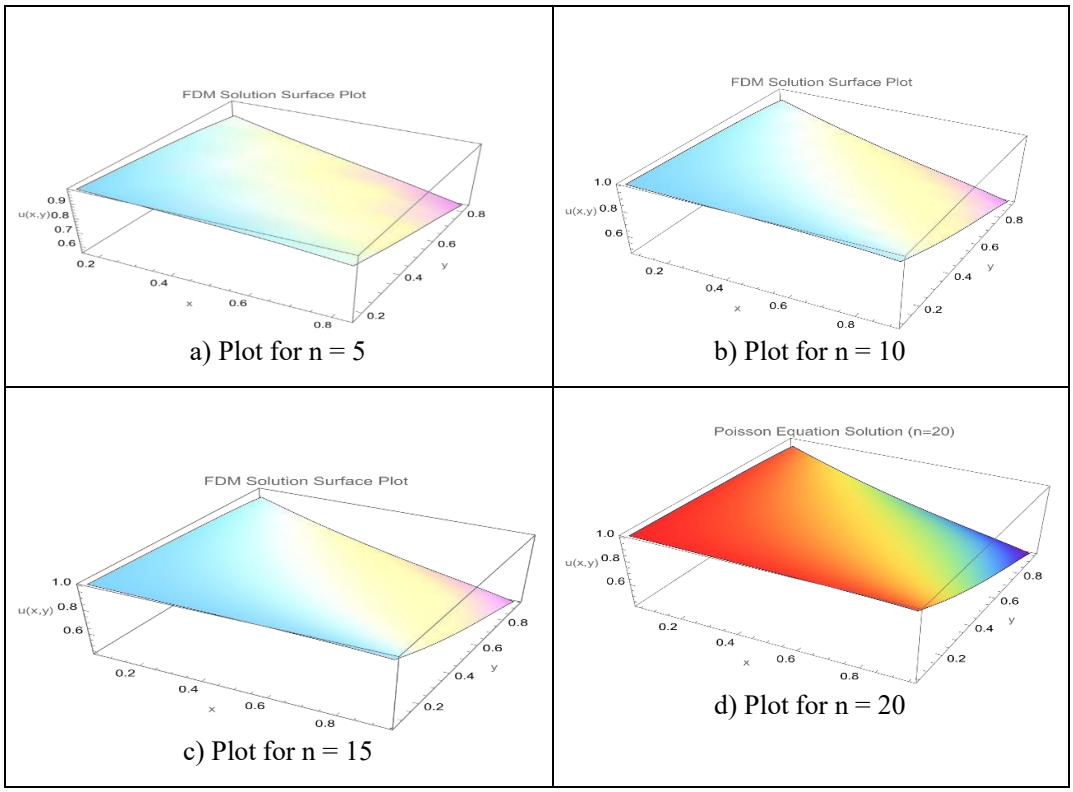
$$Absolute\ Error = |u_{exact}(x_i, y_j) - u_{numerical}(x_i, y_j)| \quad (4)$$

The convergence and accuracy were evaluated by comparing these errors across grid sizes.

4. Results and Discussion

4.1. Poisson's Equation Results

As the grid size increased, the numerical solution became smoother and more closely resembleds the exact solution. This visually confirms the accuracy improvement with finer discretization.



a) Plot for n = 5

b) Plot for n = 10

c) Plot for n = 15

d) Plot for n = 20

Figure 1: Surface plot (Poisson's Equation) for n = 5, 10, 15, 20

To quantitatively verify this, absolute errors were calculated for each case.

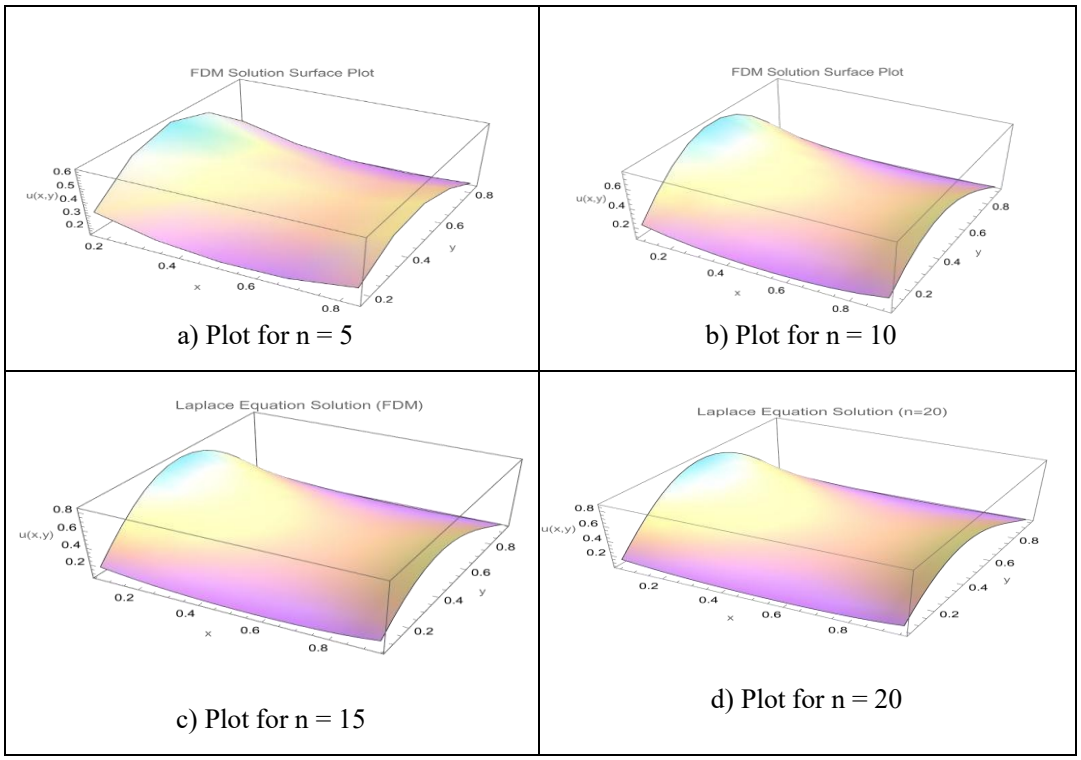
Table 1: Absolute Error (Poisson's Equation)

n	Max Error
5	0.0088
10	0.0096
15	0.0386
20	0.0097

This confirms that the five-point stencil scheme provides second-order accuracy.

4.2. Laplace's Equation Results

Although Laplace's equation in this study lacks an exact solution for the given boundary values, the numerical results were visualized using surface plots. The solution became smoother and more stable as the grid was refined. The method proves to be stable for homogeneous PDEs, with results that were consistent across different n values.



a) Plot for n = 5

b) Plot for n = 10

c) Plot for n = 15

d) Plot for n = 20

Figure 2: Surface plot (Laplace's Equation) for n = 5, 10, 15, 20

An absolute error analysis was performed by comparing the numerical results to a manufactured "exact" profile or known benchmark solution.

Table 2: Absolute Error (Laplace's Equation)

n	Max Error
5	0.0088
10	0.0096
15	0.4831
20	0.1067

Despite some local variations, the overall trend shows that numerical accuracy improved with grid refinement. The surface plots and error data collectively confirm that the FDM is a reliable and consistent method for solving Laplace's equation in structured domains.

In summary, the results verify that the Finite Difference Method combined with Dirichlet boundary conditions yields stable and accurate solutions for Laplace's equation. The absence of a

The source term did not affect the convergence behaviour, and the method proved robust under all tested grid configurations.

5. Conclusion

This project applied the Finite Difference Method (FDM) to solve Poisson's and Laplace equations under Dirichlet boundary conditions using a five-point stencil scheme. The results show that the method is accurate and stable, especially as the grid is refined. The convergence behaviour confirms the second-order accuracy of the stencil. This study demonstrates that FDM is a practical and effective technique for solving boundary value problems numerically.

Acknowledgment

The authors would like to express their deepest gratitude to Universiti Teknologi MARA (UiTM) for the opportunity to conduct this research project. Special thanks are extended to our supervisor, Dr. Azhar Bin Ahmad, and co-supervisor, Pn. Sharifah Sarimah for their continuous guidance, support, and valuable insights throughout the completion of this study.

References

Briggs, W. L., Henson, V. E., & McCormick, S. F. (2000). *A Multigrid Tutorial* (2nd ed.). Society for Industrial and Applied Mathematics (SIAM).

Fornberg, B. (1998). *A Practical Guide to Pseudospectral Methods*. Cambridge University Press.

Quarteroni, A., & Valli, A. (2008). *Numerical Approximation of Partial Differential Equations*. Springer.

Richardson, L. F. (1928). The approximate arithmetical solution by finite differences of physical problems. *Philosophical Transactions of the Royal Society A*, 226, 299–361.

Thomas, J. W. (1995). *Numerical Partial Differential Equations: Finite Difference Methods*. Springer.

Van Linde, H. J. (1974). High-order finite-difference methods for Poisson's equation. *Mathematics of Computation*, 28(126), 369–391.

Zaman, M. (2022). Numerical solution of the Poisson equation using finite difference matrix operators. *Electronics*, 11(15), 235

Zhang, K., Wang, L., & Zhang, Y. (2016). Improved finite difference method with a compact correction term. *Numerical Heat Transfer, Part B*, 70(6), 393–405.

Comparative Study of AHP and Potential Method in Supplier Selection Decision Making

Sarah Afrina Saidi[1], Wan Nur Haryani Wan Mohd Ariff Rashiddin[1], Siti Salwana Mamat[2]

[1]*Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, 70300 Seremban, Negeri Sembilan, Malaysia*

[2]*Centre of Foundation Studies, Universiti Teknologi MARA, Cawangan Selangor, Kampus Dengkil, 43800 Dengkil, Selangor, Malaysia*

[*]sitisalwana@uitm.edu.my

Abstract

Multi-Criteria Decision Making (MCDM) is used to tackle decision-making problem by evaluating and comparing different options based on several criteria. MCDM procedures need multiple inputs and generate outcomes as weights given to each options. Two commonly used MCDM methods are the Analytic Hierarchy Process (AHP) and the Potential Method (PM). AHP entails breaking down a decision problem into structured hierarchy, facilitating pairwise comparisons to determine weights for the options. On the other hand, PM utilizes a directed graph referred to as a preference structure, constructed via paired comparisons, to formulate a value function for the option. This study employs the Potential Method in a previously published case study of supplier selection that was analyzed earlier using the AHP. The objective is to analyse the outcomes from both methods, specifically focusing on the supplier's rankings. The findings showed that both AHP and PM recognized Supplier D1 as the top option, followed by Supplier D2 and then D3. The overall ranking remained consistent despite minor variations in the final weight values. This suggests that the PM produce results similar to those of AHP in tackling MCDM problems. Additionally, PM enhanced interpretability via its graphical representation, which make it a useful complementary tool to the conventional AHP for managing complex decision-making situations.

Keywords*:* Decision making, Multicriteria decision making, Analytical hierarchy process, Potential method, Supplier selection

Introduction

In today's world, people and organizations often encounter various decision making situations that involve several criteria. Although decisions can be made instinctively or without systematic approach, decision making techniques provide organized and trustworthy guidance for intricate assessments [1-2]. Multicriteria decision making (MCDM) has become essential in daily operations within numerous organizations. MCDM allow people to evaluate multiple factor at once, including cost, time, quality and risk [3].

This study emphasis on two commonly utilized MCDM techniques which are the Analytic Hierarchy Process (AHP) and the Potential Method (PM). The study focuses on improving the analysis of AHP matrices by utilizing the PM in a supplier selection case study that was evaluated earlier using AHP. The aim is to evaluate the decision results of both technique, especially regarding the resultant weights and rankings.

AHP was developed by Thomas L. Saaty during the 1970s, is a recognized technique for multicriteria decision making. It aids in decision making through pairwise comparisons to access the relative importance of different factors, while also verifying the consistency of

judgement [4-5]. Conversely, the PM is based on network analysis and effectively reveals inconsistencies in decision making data. PM works well even when the information is incomplete, concentrating on the relationships and preferences between options [6].

Literature Review

The PM is a method used in decision-making processes that make use of graphs. A structure created by comparison of a group of options or alternatives is called a preference graph. Preference graphs were employed by simulating pairwise comparisons of alternatives.

Assume that V is a collection of options or alternatives where some preferences are considered. A directed edge from vertex v to vertex u can be used to indicate that u is favoured over v (represented as $u \succ v$). The edge is denoted as (u,v).

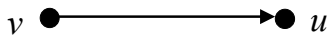


Figure 1: An alternative u is more preferred than alternative v

The PM converts pairwise comparison of options or alternatives into preference flow within a directed graph called preference graph. The preference graph contains vertex as option or alternative, and directed edge indicate preference between them. Each edge is assigned a weight that reflect to the strength of the preference. The definition of preference graph is given as follows:

Definition 1 [6]

A preference graph can be represented as a triple $G=(V,E,F)$, where $E \subseteq V \times V$ is a set of $m \in \mathbb{N}$ directed edges, V is a set of $n \in \mathbb{N}$ vertices and are representing alternatives, and $F : E \to \mathbb{R}$ is a preference flow that maps each edge (u,v) to the associated intensity $F(u,v)$.

Example 1

Assuming the preferences over A, B and C as follows:

- A to B with a weight of 3 indicates that A is preferred over B.
- B to C with a weight of 2 indicates that B is preferred over C.
- A to C with a weight of 5 indicates that A is preferred over C.

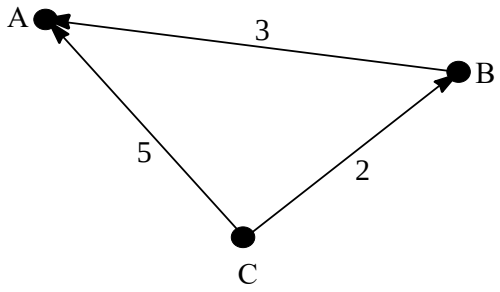


Figure 2: An example of Preference Graph of Example 1

The PM determines the 'potential' values of each alternative by balancing the incoming and outgoing preference flows at each vertex. The potential values are used to rank the options or alternatives, where higher potentials signify more preferred or desirable options. In summary, the PM employs graph theory and pairwise comparison to systematically evaluate and rank alternatives.

Ranking by Potential Method

The inputs are the preference value obtained by pairwise comparison [7]. These values are presented as the entries for a preference flow matrix. The edge's orientation is listed as an incidence matrix, A. The procedure of ranking with the PM is outlined as follows:

Step 1: For a given problem construct $G=(V,E,F)$. The V is a set of vertices, E is the set of edges and F is the corresponding preference flow.

Step 2: Construct the incidence, A and preference flow, F matrices. An $m\times n$ incident matrix is given

$$A_{\alpha,v}=\begin{cases}-1, & \text{if the edge } \alpha \text{ leaves } v\\ 1, & \text{if the edge } \alpha \text{ enters } v\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

The preference flow, F of size $m\times 1$ is the associated intensity value.

Step 3: Build the laplacian matrix, L. The laplacian matrix is given by $L=A^TA$ and its entries are defined as:

$$L_{i,j}=\begin{cases}-1, & \text{if there is an edge } (i,j) \text{ or } (j,i),\\ \deg(i), & \text{if } i=j,\\ 0, & \text{else.}\end{cases} \tag{2}$$

Step 4: Compute the flow difference matrix D. Let $D=A^TF$ be the flow difference. The following is the component of D.

$$D_v = \sum_{v=1}^{m} A_{v,\alpha}^T F_\alpha = \sum_{\alpha \text{ enters } v} F_\alpha - \sum_{\alpha \text{ leaves } v} F_\alpha \tag{3}$$

In this case, D represents the difference between the total flow entering v and the total flow leaving v.

Step 5: Solve for potentials. Potential, X is a Laplacian system solution.

$$LX = \nabla \tag{4}$$

In a way, that $\sum X_v = 0$ on its linked parts.

Step 6: Check consistency degree $\beta < 12^\circ$. The definition of the measure of inconsistency is

$$\text{Inc}(F) = \frac{\|F - AX\|_2}{\|AX\|_2} \tag{5}$$

where $\beta = \arctan\ \text{Inc}(F)$ is the angle of inconsistency and $\|\ .\ \|_2$ indicates the 2-norm. When $\beta < 12^\circ$, the ranking is acceptable.

Step 7: Calculate weights, w. The weight is calculated using the following formula.

$$w = \frac{a^X}{\|a^X\|_1}, \quad a > 0 \tag{6}$$

where the parameter a is set to 2 and $\|\ .\ \|_1$ stands for the l_1-norm.

Step 8: Rank the items according to the corresponding weights.

Implementation on supplier selection case study

Czech manufacturing company is needed to find the best supplier [8]. The company is seeking a new supplier for a particular item. There are 10 criteria to be evaluated to find the best suppliers. The company has three potential suppliers and is using a structured, multi-criteria decision-making approach to evaluate them.

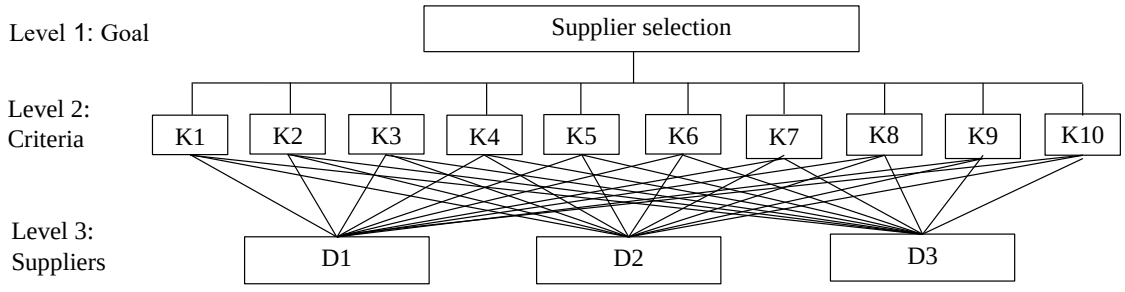


Figure 3: Hierarchy of problem.

Criteria weight determination

The initial step in analyzing hierarchical decision problem is find the weight of criteria. Table 1 show the comparison matrix of 10 criteria taken form [8]. Table 1 represent the pairwise comparison of criteria in consideration to selection of suppliers and the combination of weight and ranking for criteria which is measured AHP and PM.

Table 1: Pairwise comparison matrix of criteria.

Criteria	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
K1	1	3	5	5	7	8	8	6	5	5
K2	1/3	1	5	5	7	8	8	6	5	5
K3	1/5	1/5	1	1/3	5	6	7	5	3	4
K4	1/5	1/5	3	1	4	3	7	5	1	5
K5	1/7	1/7	1/5	1/4	1	4	1/5	1/7	1/3	3
K6	1/8	1/8	1/6	1/3	1/4	1	5	1/5	1/3	3
K7	1/8	1/8	1/7	1/7	5	1/5	1	1/3	1/3	3
K8	1/6	1/6	1/5	1/5	7	5	3	1	1	3
K9	1/5	1/5	1/3	1	3	3	3	1	1	3
K10	1/5	1/5	1/4	1/5	1/3	1/3	1/3	1/3	1/3	1

The pairwise comparisons among 10 criteria are transformed into a preference graph as in Figure 4.

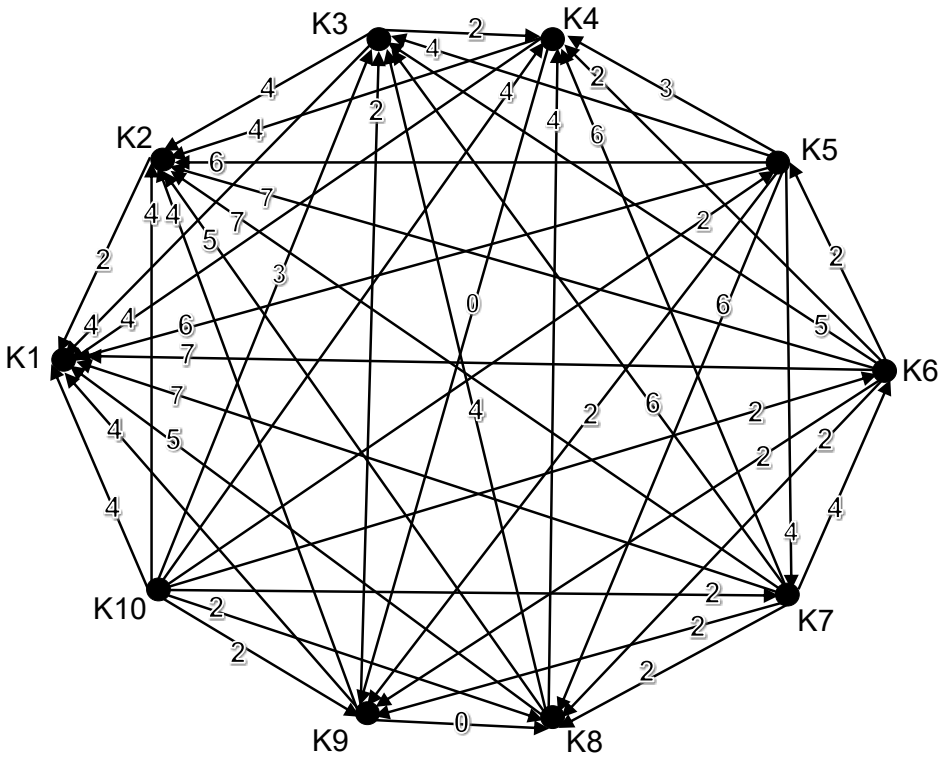


Figure 4: The preference graph of criteria.

The incident matrix, A of size 45 x 10 is constructed using Equation (1). However, only the elements in the first row of the transpose incident matrix, A^T provided as follows:

$$\text{Row}(A^T)=\left[A_{1j}\right]_{j=1}^{45}$$
$$=[1\,1\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0

Supplier weight with respect to K1

Once the weights for the criteria are obtained, the third level is analyzed. This section discusses the analysis of suppliers with respect to criterion, K1. Table 3 presents the pairwise comparison of suppliers based on K1. The information from Table 3 are then transformed into preference graph, as illustrated is Figure 5. In this graph, the directed edges from D3 to A2 signifies that D2 is preferred over D3 by a factor of 4.

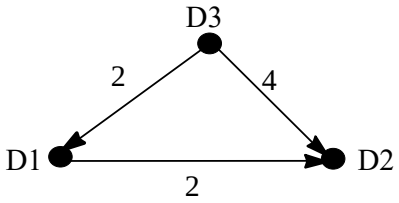


Figure 5: The preference graph of suppliers with respect to K1.

Table 3 shows the comparative weights of suppliers based on criterion K1, as determined by both AHP and PM. The PM ranks the suppliers in the order D2>D1>D3, which consistent with the ranking obtained from AHP.

Table 3: Weightage for supplier with respect to K1.

Suppliers	D1	D2	D3	AHP		PM	
				weight	rank	weight	rank
D1	1	1/3	3	0.2583	2	0.1905	2
D2	3	1	5	0.6370	1	0.7619	1
D3	1/3	1/5	1	0.1047	3	0.0476	3

Conclusion

Table 4 shows the comparison of final weight and ranking between the AHP and PM. The AHP find outs that D1 holds the highest priority with a weight of (0.4785), followed by D2 (0.4397), while D3 has the lowest weight which is 0.0816. Similarly, the PM analysis ranks D1 highest (0.5753), then D2 (0.4050), and lastly D3 (0.0197). This alignment in ranking across both methods confirms a consistent order of alternatives as D1 > D2 > D3.

Table 4: Final weights of suppliers using AHP and PM.

Suppliers	AHP		PM	
	weight	rank	weight	rank
D1	0.4785	1	0.5753	1
D2	0.4397	2	0.4050	2
D3	0.0816	3	0.0197	3

In conclusion, both AHP and PM rank the supplier identically with D1 first, D2 second and D3 third. The final weight values between these two methods are differed slightly, but the order

of preference was the same. This shows that PM is stable and consistent, thus comparable to the AHP.

Acknowledgments

The authors would like to express their sincere appreciation to Universiti Teknologi MARA, Seremban Campus, for its continuous support. Special thanks are also extended to the reviewers for their constructive comments and insightful recommendations.

References

[1] Taherdoost, H., & Madanchian, M. (2024). Decision making: Models, processes, techniques. *Cloud Computing and Data Science*, 1-14.

[2] U-Dominic, C. M., Ujam, J. C., & Igbokwe, N. (2021). Applications of analytical hierarchy process (AHP) and knowledge management (km) concepts in defect identification: a case of cable manufacturing. *Asian Journal of Advanced Research and Reports*, 9-21.

[3] Taherdoost, H., & Madanchian, M. (2023). Multi-criteria decision making (MCDM) methods and concepts. *Encyclopedia*, 3(1), 77-87.

[4] Huang, J. J. (2021). Analytic Hierarchy Process with The Correlation Effect Via Wordnet. *Mathematics*, 9(8), 872. https://doi.org/10.3390/math9080872.

[5] Schiavon, L. L. P., Lima, P. A. B., Crepaldi, A. F., & Mariano, E. B. (2023). Use of The Analytic Hierarchy Process Method in The Variety Selection Process for Sugarcane Planting. *Eng*, 4(1), 602–614. https://doi.org/10.3390/eng4010036

[6] Čaklović, L., & Kurdija, A. S. (2017). A Universal Voting System Based on the Potential Method. *European Journal of Operational Research*, 259(2), 677–688. https://doi.org/10.1016/J.EJOR.2016.10.032

[7] Johar, K. A., Mohd Amin, M. A. H. N., Tuah, M. F., & Mamat, S. S. (2025). The graphical representation of clothing store location and its application with potential method. *Journal of Administrative Science*, 22, 34-46.

[8] Hruška, R., Průša, P., & Babić, D. (2014). The Use of AHP Method for Selection of Supplier. *Transport*, 29(2), 195–203. https://doi.org/10.3846/16484142.2014.930928

APPROXIMATE SOLUTION OF FRACTIONAL DIFFERENTIAL EQUATIONS USING TRANSFORMATION AND PERTURBATION METHOD

Hana Zaidah Zulkefli [1] and Farah Suraya Md Nasrudin [2,*]
[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, [2]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Johor Branch, Segamat Campus
[*]farah53@uitm.edu.my

Abstract

Fractional calculus extends basic calculus, specifically differentiation and integration to non-integer orders which include real and complex numbers, enabling more accurate modelling of actual systems. The fractional differential equation (FDEs) is one kind of fractional calculus that has been widely implemented in a variety of fields, including biology, engineering, and physics. FDE extends ordinary and partial differential equations of non-integer order. This makes them more difficult to solve, since there are very few well-developed analytical methods available to solve FDE that have a complicated solution. This study explores the numerical methods, the Homotopy Perturbation Transform Method (HPTM) in solving FDE, specifically time-fractional diffusion equations, using the Riemann-Liouville definition of fractional derivatives. HPTM combines the Laplace transform as well as Homotopy Perturbation Method (HPM) that simplifies the equation and approaches the solution efficiently. Thus, the findings indicate that the suggested method is highly efficient and is able to simplify the solution of FDE. The solutions were also analyzed using Maple software and have been plotted for various fractional derivative orders, α and several numerical illustrations are provided.

Keywords: fractional derivative equation, fractional diffusion equation, homotopy perturbation method, laplace transform

1. Introduction

Fractional calculus is a generalization of basic calculus, specifically in differentiation and integration (Fernandez and Fahad, 2022). The theory covers differentiation and integration functions to non-integer orders, which include real and complex numbers (Ma et al., 2018). This is different from classical calculus which is only limited to integer order of derivatives and integrals. The theory of fractional calculus has existed since 1695. Its origins can be traced back to the interaction between L'Hôpital and Leibniz over the non-integer order of derivatives (Munoz-Pacheco et al., 2023). Fractional calculus can be applied in various areas (Khan et al., 2012). This makes it a valuable tool for researchers exploring systems with non-integer characteristics.

Fractional differential equations (FDE) belong to fractional calculus, which extends ordinary and partial differential equations to non-integer orders. This kind of equation is needed in various complex systems in real life. It is important in modelling nonlinear mechanisms and structures in various fields (Khan et al., 2012). For example, in medicine, biology, chemistry, mathematics, physics, engineering, and economics (including control theory and the oil industry) (Hailat et al., 2019). FDE is also a useful tool for characterizing long-memory processes and materials, anomalous diffusion, long-range interactions, long-term behaviour, power laws, and even allometric scaling laws (Zheng et al., 2019).

When referring to the fractional derivatives, several mathematicians approach this problem from different perspectives and provide various definitions. It includes fractional derivatives of Riemann-Liouville, Caputo, Hadamard, Weyl, and Riesz (Oliveira and Machado, 2014). This study will focus on fractional derivative in the Riemann-Liouville approach.

It is important to highlight that the FDE structure is more challenging to solve than the classical differential equation and is also very difficult to solve analytically. This because FDE solutions, which have non-integer order, may not have simple closed-form expressions. Therefore, numerical solutions are usually used to approximate the solution of this equation. There are many numerical methods mentioned in past research, among them are homotopy perturbation method (HPM), the variation iteration method

(VIM), the Adomian decomposition method (ADM), and the Bessel collocation method (BCM) (Hailat et al., 2019).

This study will focus on the homotopy perturbation transform method (HPTM) which refers to the combination of the Laplace transform and the homotopy perturbation method (HPM). HPTM will be used to obtain an approximate solution for FDE. Additionally, time-fractional diffusion equation is a form of FDE that describes the free movement of particles modelled by the classical diffusion equation (Kumar et al., 2012). Particularly, in this study, HPTM will be used to solve the time-fractional diffusion equations in the Riemann-Liouville sense. The general form of the time-fractional diffusion equation is

$$\frac{\partial^\alpha u}{\partial t^\alpha} = \psi \frac{\partial u}{\partial x}, \tag{1}$$

where ψ is the diffusion coefficient. In fractional diffusion, the order of the derivative, α is a non-integer, allowing for more flexibility in modeling anomalous diffusion phenomena.

The structure of this article is as follows. In the following sections, we will briefly introduce a few preliminary concepts, such as the homotopy perturbation method (HPM), the Laplace transform and the Riemann-Liouville derivative. The details of the proposed method are presented in a subsequent section. Following that, some examples are provided to demonstrate the application of the suggested method in solving fractional-order diffusion equations using the Riemann-Liouville derivative. We also presented the numerical experiments that show the validity of the proposed method. The final section emphasises the conclusion and we made some recommendations.

2. Preliminaries

We will concentrate on the following time-fractional order diffusion equation:

$$\frac{\partial^\alpha u}{\partial t^\alpha} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial}{\partial x}(F(x)u), \tag{2}$$

where the fractional derivative term is in the Riemann-Liouville derivative sense. The parameter α refers to order of the time-fractional derivative term with $0 < \alpha < 1$.

2.1. *Riemann-Liouville derivative*

In this section, we will briefly present some basic definitions related to the Riemann-Liouville fractional derivative.

Definition 2.1. The Riemann-Liouville fractional derivative is defines as

$${}^{RL}D_x^\alpha f(x) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dx} \int_a^x \frac{f(t)}{(x-t)^\alpha} dt \tag{3}$$

Equation (3) is the most widely known definition of the fractional derivative; it is usually called the Riemann–Liouville definition.

Definition 2.2. The Laplace transform of the Riemann-Liouville fractional derivative is defines as

$$\mathscr{L}\{D_x^\alpha f(x)\} = s^\alpha F(s) - \sum_{k=0}^{n-1} s^k D_x^{\alpha-k-1} f(0) \tag{4}$$

where $n-1 < \alpha \leq n$.

2.2. *Laplace transform*

Laplace transform will be applied to simplify the equation in the first phase. The following steps will describe how to apply Laplace transform to simplify the FDE, and most of these steps are taken from Kumar et al. (2012).

For an overview of the HPTM procedure, the following is a nonlinear fractional differential equation:

$$D_t^{\alpha} u(x,t) + R(x)u(x,t) + N(x)u(x,t) = q(x,t), \quad t > 0, \tag{5}$$

where N is the nonlinear operator, R is the linear operator and $q(x,t)$ is continuous function.

Laplace transform will be applied to both sides of (5):

$$\mathscr{L}\{D_t^{\alpha} u(x,t)\} + \mathscr{L}\{R(x)u(x,t) + N(x)u(x,t)\} = \mathscr{L}\{q(x,t)\}. \tag{6}$$

The derivative property of the Laplace transform (4) is applied in (6)

$$\mathscr{L}\{u(x,t)\} = s^{-\alpha} D_t^{\alpha-1} u(x,0) + s^{-\alpha}\mathscr{L}\{q(x,t)\} - s^{-\alpha}\mathscr{L}\{R(x)u(x,t) + N(x)u(x,t)\}. \tag{7}$$

The inverse Laplace transform is then applied to both sides to get the solution

$$u(x,t) = G(x,t) - \mathscr{L}^{-1}\big\{s^{-\alpha}\mathscr{L}\{R(x)u(x,t) + N(x)u(x,t)\}\big\}, \tag{8}$$

where $G(x,t)$ defined the term arising form the source term and the prescribed initial conditions.

2.3. *Homotopy perturbation method*

Homotopy perturbation method (HPM) is one of the analytical approaches to estimate the solution of FDE (Khan et al., 2012). HPM is needed to obtain efficient analytical and numerical methods since it is known for its efficiency and simplicity (Javeed et al., 2019).

In the second phase, HPM will be implemented to approximate the solution. The solution can be expressed as a power series in p:

$$u(x,t) = \sum_{n=0}^{\infty} p^n u_n(x,t), \tag{9}$$

where the homotopy parameter, p is considered as a small parameter.

The nonlinear term can be expressed as

$$Nu(x,t) = \sum_{n=0}^{\infty} p^n H_n(u), \tag{10}$$

where H_n is the He's polynomial

$$H_n(u_0, u_1, u_2, \ldots, u_n) = \frac{1}{n!}\frac{\partial^n}{\partial p^n}\left[N\left(\sum_{i=0}^{\infty} p^i u_i\right)\right], \; n = 0, 1, 2, 3, \ldots. \tag{11}$$

Substituting (9), (10) and (11) into (8) we get

$$\sum_{n=0}^{\infty} p^n u_n(x,t) = G(x,t) - p\mathscr{L}^{-1}\big\{s^{-\alpha}\mathscr{L}\{R\sum_{n=0}^{\infty} p^n u_n(x,t) + N\sum_{n=0}^{\infty} p^n H_n(u)\}\big\} \tag{12}$$

Equating terms to the same power of p on both sides will get

$$p^0 : \; u_0(x,t) = G(x,t) \tag{13}$$

$$p^n : \; u_n(x,t) = \mathscr{L}^{-1}\big\{s^{-\alpha}\mathscr{L}\{Ru_{n-1}(x,t) + H_{n-1}(u)\}\big\}, \; n > 0. \tag{14}$$

3. Homotopy Perturbation Transform Method

This Homotopy perturbation transformation method (HPTM) is a method which combines two specific methods, the Laplace transform and the homotopy perturbation method (HPM). Laplace transform is applied to simplify the equation. Then, we implement the homotopy perturbation method to approximate the solution of fractional differential equations (FDE).

There are three phases in solving fractional differential equations (FDE). The first phase is to apply the Laplace transform to simplify the equation. The second phase is to apply homotopy perturbation method (HPM) to approximate the solution. The next section explains how HPTM is used to solve the equations.

4. Numerical Examples

The purpose of using Homotopy perturbation method (HPTM) to solve FDE is to prove its simplicity and efficiency. In order to verify the efficiency of the proposed method, a comparison between the approximate solution obtained and the exact solution is carried out. Approximate solutions will be evaluated at different values of α. Moreover, the results of this study will also be analysed using Maple software. Therefore, in this article, apart from manual calculations, Maple software will also be used to test the efficiency of HPTM by comparing the approximate solutions at different α values with the exact solutions.

There are two examples to demonstrate how HPTM was being applied to solve time-fractional diffusion equations.

Example 4.1. Consider the following time-fractional diffusion equation

$$\frac{\partial^\alpha u(x,t)}{\partial t^\alpha} = \frac{\partial^2 u(x,t)}{\partial x^2} + \frac{\partial}{\partial x} x u(x,t) \tag{15}$$

with the initial condition $u(x,0) = 1$.

Applying the steps as explained in Section 2 and using $\alpha = 0.5$, we get

$$u(x,t) = \frac{2}{\pi} + \mathscr{L}^{-1}\{s^{-1/2}\mathscr{L}(D_x^2 u(x,t) + D_x x u(x,t))\}. \tag{16}$$

Substituting (12) into (16) and equating terms to the same order of p we get

$$\begin{aligned}
p^0 &: \ u_0(x,t) = \frac{2}{\pi} \\
p^1 &: \ u_1(x,t) = \frac{4\sqrt{t}}{\pi^{3/2}} \\
p^2 &: \ u_2(x,t) = \frac{2t}{\pi} \\
p^3 &: \ u_3(x,t) = \frac{t^2}{\pi} \\
&\vdots
\end{aligned}$$

As shown above, the sample solution for Example 4.1 with $\alpha = 0.5$ by using HPTM has been achieved. We also show our numerical result graphically in Figure 1 for the comparison of approximation between different values of fractional order.

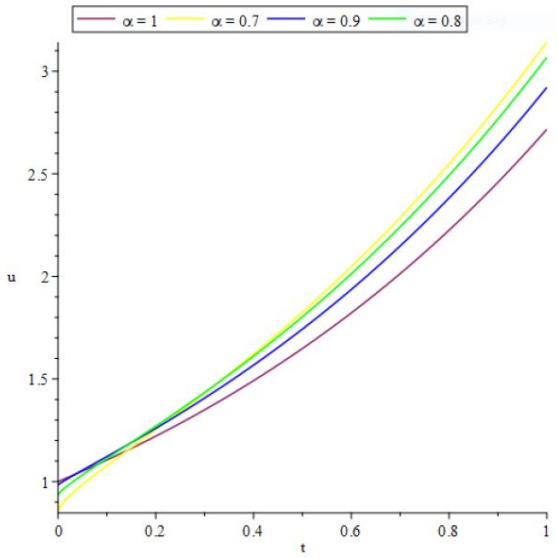


Figure 1: Comparison of solution $u(x,t)$ at different values of α in Example 4.1

From Figure 4.1, it is observed that as α decreases, the solution $u(x,t)$ increases for the same values of x and t. This indicated that when the value of $\alpha < 1$, the equation incorporates fractional effects, which increase $u(x,t)$.

Example 4.2. Consider the following time-fractional diffusion equation

$$\frac{\partial^\alpha u(x,y,t)}{\partial t^\alpha} = \frac{\partial^2 u(x,y,t)}{\partial x^2} + \frac{\partial^2 u(x,y,t)}{\partial y^2} \tag{17}$$

with the initial condition $u(x,y,0) = \sin x \sin y$.

Applying the steps as explained in Section 2 and using $\alpha = 0.5$, we get

$$u(x,y,t) = \frac{2\sin x \sin y}{\pi} + \mathscr{L}^{-1}\{s^{-1/2}\mathscr{L}(D_x^2 u(x,t) + D_y^2 u(x,t))\}. \tag{18}$$

Substituting (12) into (18) and equating terms to the same order of p we get

$$\begin{aligned}
p^0 &: \ u_0(x,t) = \frac{2\sin x \sin y}{\pi} \\
p^1 &: \ u_1(x,t) = \frac{-8\sin x \sin y \sqrt{t}}{\pi^{3/2}} \\
p^2 &: \ u_2(x,t) = \frac{8t \sin x \sin y}{\pi} \\
p^3 &: \ u_3(x,t) = \frac{-64t^2 \sin x \sin y}{3\pi^{3/2}} \\
&\vdots
\end{aligned}$$

The calculation above is the sample solution for Example 4.2 when $\alpha = 0.5$ by using HPTM. The comparison of approximation between different values of fractional order, α is presented graphically in Figure 2 below.

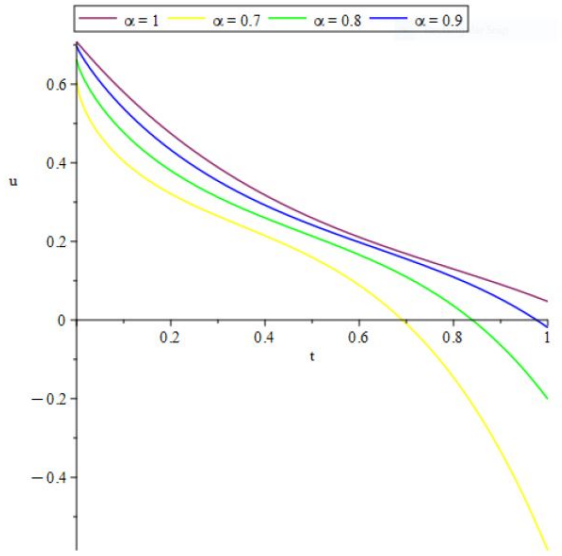


Figure 2: Comparison of solution $u(x,t)$ at different values of α in Example 4.2

From Figure 2, the solution obtained for Example 4.2 with the HPTM method appears to decrease very quickly with the increase in α at the value of $x = y = 1$. We can see a similar behaviour in the approximate solution, $u(x,y,t)$ for this example, same as in Example 4.1 when α has evolved.

Figure 3 illustrates the comparative solution between the exact solution and the approximate solution using HPTM for $\alpha = 1$.

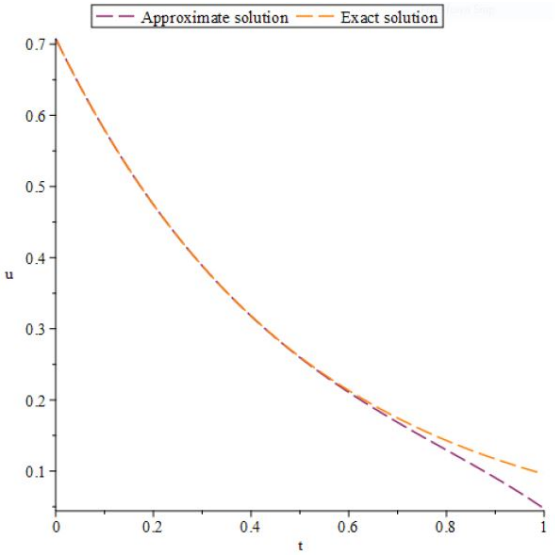


Figure 3: Comparison of exact solution and approximate solution for $\alpha = 1$

Here, we compare our solution, (18) with the exact solution, i.e. $u(x,t) = e^{-2t}\sin x \sin y$ having $\alpha = 1$ (Kumar et al., 2012). From the numerical result, we obtained a good result in comparison to the exact solution.

5. Conclusion

This study's main contribution is implementing a method to solve FDE in the Riemann-Liouville sense. In this section, the convergence of the FDE solution can be shown by using HPTM, and it proves that this method is easy to use than other methods. The efficiency of the FDE solution obtained by using HPTM with different values of α. The graph shown in Figure 1 and Figure 2 is a comparison graph of the approximate solution for different values of $\alpha = 0.7, 0.8, 0.9, 1$. When $\alpha = 1$, the equation corresponds to a standard differential equation of integer order. As $\alpha < 1$, the equation incorporates fractional derivative effects, which increase or decrease $u(x, t)$. This behaviour in fractional derivative equations shows that the order of derivative, α governs the influence of past states or memory effects on the current states of the equation.

The main goal of this study was to apply the Homotopy Perturbation Transformation Method (HPTM) to find approximate solutions for fractional differential equations (FDE). The achievement of this objective was eased by the application of the method that based on the Riemann-Liouville derivative on a specific class of FDE.

The selection of HPTM resulted from its unique feature: it balances analytical and numerical efficiency. HPTM is considered the convenient method for dealing with FDE problems. This method offers the advantage of a simpler calculation steps when compared to several other numerical and analytical techniques. The Variational Iteration Method (VIM) provides a particularly useful example for comparison. Selecting an appropriate multiplier for the mathematical problem is important in the VIM. This selection process can necessitate meaningful manual adjustment of the initial conditions. This study reveals the important potential of HPTM as an effective alternative for solving FDE, thus contributing to at least some progress in several future applications within the field of FDE.

References

Fernandez, A. and Fahad, H. M. (2022). Weighted fractional calculus: a general class of operators. *Fractal and Fractional*, 6(4):208.

Hailat, I., Nemrat, A. A., Zainuddin, Z., and Azmi, A. B. (2019). Analytical solution of ordinary fractional differential equations by modified homotopy perturbation method and laplace transform. In *Journal of Physics: Conference Series*, volume 1366, page 012037. IOP Publishing.

Javeed, S., Baleanu, D., Waheed, A., Shaukat Khan, M., and Affan, H. (2019). Analysis of homotopy perturbation method for solving fractional order differential equations. *Mathematics*, 7(1):40.

Khan, Y., Faraz, N., Kumar, S., and Yildirim, A. (2012). A coupling method of homotopy perturbation and laplace transformation for fractional models. *University" Politehnica" of Bucharest Scientific Bulletin, Series A: Applied Mathematics and Physics*, 74(1):57–68.

Kumar, S., Yildirim, A., Khan, Y., and Wei, L. (2012). A fractional model of the diffusion equation and its analytical solution using laplace transform. *Scientia Iranica*, 19(4):1117–1123.

Ma, Q., Lin, X., and Li, H. (2018). A comparative teaching of fractional calculus and integer calculus. In *2018 International Conference on Social Science and Education Reform (ICSSER 2018)*, pages 94–97. Atlantis Press.

Munoz-Pacheco, J. M., Wei, Z., Volos, C., and Sambas, A. (2023). Future challenges in the fractional-order dynamical systems: from mathematics to applications.

Oliveira, E. C. d. and Machado, J. (2014). A review of definitions for fractional derivatives and integral. *Mathematical Problems in Engineering*, 2014:1–7.

Zheng, Z., Zhao, W., and Dai, H. (2019). A new definition of fractional derivative. *International Journal of Non-Linear Mechanics*, 108:1–6.

APPROXIMATE SOLUTION OF FRACTIONAL DIFFERENTIAL EQUATIONS USING TRANSFORMATION AND PERTURBATION METHOD

Hana Zaidah Zulkefli [1] and Farah Suraya Md Nasrudin [2,*]
[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, [2]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Johor Branch, Segamat Campus
*farah53@uitm.edu.my

Abstract

Fractional calculus extends basic calculus, specifically differentiation and integration to non-integer orders which include real and complex numbers, enabling more accurate modelling of actual systems. The fractional differential equation (FDEs) is one kind of fractional calculus that has been widely implemented in a variety of fields, including biology, engineering, and physics. FDE extends ordinary and partial differential equations of non-integer order. This makes them more difficult to solve, since there are very few well-developed analytical methods available to solve FDE that have a complicated solution. This study explores the numerical methods, the Homotopy Perturbation Transform Method (HPTM) in solving FDE, specifically time-fractional diffusion equations, using the Riemann-Liouville definition of fractional derivatives. HPTM combines the Laplace transform as well as Homotopy Perturbation Method (HPM) that simplifies the equation and approaches the solution efficiently. Thus, the findings indicate that the suggested method is highly efficient and is able to simplify the solution of FDE. The solutions were also analyzed using Maple software and have been plotted for various fractional derivative orders, α and several numerical illustrations are provided.

Keywords: fractional derivative equation, fractional diffusion equation, homotopy perturbation method, laplace transform

1. Introduction

Fractional calculus is a generalization of basic calculus, specifically in differentiation and integration (Fernandez and Fahad, 2022). The theory covers differentiation and integration functions to non-integer orders, which include real and complex numbers (Ma et al., 2018). This is different from classical calculus which is only limited to integer order of derivatives and integrals. The theory of fractional calculus has existed since 1695. Its origins can be traced back to the interaction between L'Hôpital and Leibniz over the non-integer order of derivatives (Munoz-Pacheco et al., 2023). Fractional calculus can be applied in various areas (Khan et al., 2012). This makes it a valuable tool for researchers exploring systems with non-integer characteristics.

Fractional differential equations (FDE) belong to fractional calculus, which extends ordinary and partial differential equations to non-integer orders. This kind of equation is needed in various complex systems in real life. It is important in modelling nonlinear mechanisms and structures in various fields (Khan et al., 2012). For example, in medicine, biology, chemistry, mathematics, physics, engineering, and economics (including control theory and the oil industry) (Hailat et al., 2019). FDE is also a useful tool for characterizing long-memory processes and materials, anomalous diffusion, long-range interactions, long-term behaviour, power laws, and even allometric scaling laws (Zheng et al., 2019).

When referring to the fractional derivatives, several mathematicians approach this problem from different perspectives and provide various definitions. It includes fractional derivatives of Riemann-Liouville, Caputo, Hadamard, Weyl, and Riesz (Oliveira and Machado, 2014). This study will focus on fractional derivative in the Riemann-Liouville approach.

It is important to highlight that the FDE structure is more challenging to solve than the classical differential equation and is also very difficult to solve analytically. This because FDE solutions, which have non-integer order, may not have simple closed-form expressions. Therefore, numerical solutions are usually used to approximate the solution of this equation. There are many numerical methods mentioned in past research, among them are homotopy perturbation method (HPM), the variation iteration method

(VIM), the Adomian decomposition method (ADM), and the Bessel collocation method (BCM) (Hailat et al., 2019).

This study will focus on the homotopy perturbation transform method (HPTM) which refers to the combination of the Laplace transform and the homotopy perturbation method (HPM). HPTM will be used to obtain an approximate solution for FDE. Additionally, time-fractional diffusion equation is a form of FDE that describes the free movement of particles modelled by the classical diffusion equation (Kumar et al., 2012). Particularly, in this study, HPTM will be used to solve the time-fractional diffusion equations in the Riemann-Liouville sense. The general form of the time-fractional diffusion equation is

$$\frac{\partial^\alpha u}{\partial t^\alpha} = \psi \frac{\partial u}{\partial x}, \tag{1}$$

where ψ is the diffusion coefficient. In fractional diffusion, the order of the derivative, α is a non-integer, allowing for more flexibility in modeling anomalous diffusion phenomena.

The structure of this article is as follows. In the following sections, we will briefly introduce a few preliminary concepts, such as the homotopy perturbation method (HPM), the Laplace transform and the Riemann-Liouville derivative. The details of the proposed method are presented in a subsequent section. Following that, some examples are provided to demonstrate the application of the suggested method in solving fractional-order diffusion equations using the Riemann-Liouville derivative. We also presented the numerical experiments that show the validity of the proposed method. The final section emphasises the conclusion and we made some recommendations.

2. Preliminaries

We will concentrate on the following time-fractional order diffusion equation:

$$\frac{\partial^\alpha u}{\partial t^\alpha} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial}{\partial x}(F(x)u), \tag{2}$$

where the fractional derivative term is in the Riemann-Liouville derivative sense. The parameter α refers to order of the time-fractional derivative term with $0 < \alpha < 1$.

2.1. *Riemann-Liouville derivative*

In this section, we will briefly present some basic definitions related to the Riemann-Liouville fractional derivative.

Definition 2.1. The Riemann-Liouville fractional derivative is defines as

$${}^{RL}D_x^\alpha f(x) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dx} \int_a^x \frac{f(t)}{(x-t)^\alpha} dt \tag{3}$$

Equation (3) is the most widely known definition of the fractional derivative; it is usually called the Riemann–Liouville definition.

Definition 2.2. The Laplace transform of the Riemann-Liouville fractional derivative is defines as

$$\mathscr{L}\{D_x^\alpha f(x)\} = s^\alpha F(s) - \sum_{k=0}^{n-1} s^k D_x^{\alpha-k-1} f(0) \tag{4}$$

where $n-1 < \alpha \leq n$.

2.2. *Laplace transform*

Laplace transform will be applied to simplify the equation in the first phase. The following steps will describe how to apply Laplace transform to simplify the FDE, and most of these steps are taken from Kumar et al. (2012).

For an overview of the HPTM procedure, the following is a nonlinear fractional differential equation:

$$D_t^{\alpha} u(x,t) + R(x)u(x,t) + N(x)u(x,t) = q(x,t), \quad t > 0, \tag{5}$$

where N is the nonlinear operator, R is the linear operator and $q(x,t)$ is continuous function.

Laplace transform will be applied to both sides of (5):

$$\mathscr{L}\{D_t^{\alpha} u(x,t)\} + \mathscr{L}\{R(x)u(x,t) + N(x)u(x,t)\} = \mathscr{L}\{q(x,t)\}. \tag{6}$$

The derivative property of the Laplace transform (4) is applied in (6)

$$\mathscr{L}\{u(x,t)\} = s^{-\alpha} D_t^{\alpha-1} u(x,0) + s^{-\alpha}\mathscr{L}\{q(x,t)\} - s^{-\alpha}\mathscr{L}\{R(x)u(x,t) + N(x)u(x,t)\}. \tag{7}$$

The inverse Laplace transform is then applied to both sides to get the solution

$$u(x,t) = G(x,t) - \mathscr{L}^{-1}\big\{s^{-\alpha}\mathscr{L}\{R(x)u(x,t) + N(x)u(x,t)\}\big\}, \tag{8}$$

where $G(x,t)$ defined the term arising form the source term and the prescribed initial conditions.

2.3. *Homotopy perturbation method*

Homotopy perturbation method (HPM) is one of the analytical approaches to estimate the solution of FDE (Khan et al., 2012). HPM is needed to obtain efficient analytical and numerical methods since it is known for its efficiency and simplicity (Javeed et al., 2019).

In the second phase, HPM will be implemented to approximate the solution. The solution can be expressed as a power series in p:

$$u(x,t) = \sum_{n=0}^{\infty} p^n u_n(x,t), \tag{9}$$

where the homotopy parameter, p is considered as a small parameter.

The nonlinear term can be expressed as

$$Nu(x,t) = \sum_{n=0}^{\infty} p^n H_n(u), \tag{10}$$

where H_n is the He's polynomial

$$H_n(u_0, u_1, u_2, \ldots, u_n) = \frac{1}{n!}\frac{\partial^n}{\partial p^n}\left[N\left(\sum_{i=0}^{\infty} p^i u_i\right)\right], \; n = 0, 1, 2, 3, \ldots. \tag{11}$$

Substituting (9), (10) and (11) into (8) we get

$$\sum_{n=0}^{\infty} p^n u_n(x,t) = G(x,t) - p\mathscr{L}^{-1}\big\{s^{-\alpha}\mathscr{L}\{R\sum_{n=0}^{\infty} p^n u_n(x,t) + N\sum_{n=0}^{\infty} p^n H_n(u)\}\big\} \tag{12}$$

Equating terms to the same power of p on both sides will get

$$p^0 : \; u_0(x,t) = G(x,t) \tag{13}$$

$$p^n : \; u_n(x,t) = \mathscr{L}^{-1}\big\{s^{-\alpha}\mathscr{L}\{Ru_{n-1}(x,t) + H_{n-1}(u)\}\big\}, \; n > 0. \tag{14}$$

3. Homotopy Perturbation Transform Method

This Homotopy perturbation transformation method (HPTM) is a method which combines two specific methods, the Laplace transform and the homotopy perturbation method (HPM). Laplace transform is applied to simplify the equation. Then, we implement the homotopy perturbation method to approximate the solution of fractional differential equations (FDE).

There are three phases in solving fractional differential equations (FDE). The first phase is to apply the Laplace transform to simplify the equation. The second phase is to apply homotopy perturbation method (HPM) to approximate the solution. The next section explains how HPTM is used to solve the equations.

4. Numerical Examples

The purpose of using Homotopy perturbation method (HPTM) to solve FDE is to prove its simplicity and efficiency. In order to verify the efficiency of the proposed method, a comparison between the approximate solution obtained and the exact solution is carried out. Approximate solutions will be evaluated at different values of α. Moreover, the results of this study will also be analysed using Maple software. Therefore, in this article, apart from manual calculations, Maple software will also be used to test the efficiency of HPTM by comparing the approximate solutions at different α values with the exact solutions.

There are two examples to demonstrate how HPTM was being applied to solve time-fractional diffusion equations.

Example 4.1. Consider the following time-fractional diffusion equation

$$\frac{\partial^\alpha u(x,t)}{\partial t^\alpha} = \frac{\partial^2 u(x,t)}{\partial x^2} + \frac{\partial}{\partial x} x u(x,t) \tag{15}$$

with the initial condition $u(x,0) = 1$.

Applying the steps as explained in Section 2 and using $\alpha = 0.5$, we get

$$u(x,t) = \frac{2}{\pi} + \mathscr{L}^{-1}\{s^{-1/2}\mathscr{L}(D_x^2 u(x,t) + D_x x u(x,t))\}. \tag{16}$$

Substituting (12) into (16) and equating terms to the same order of p we get

$$\begin{aligned}
p^0 &: \ u_0(x,t) = \frac{2}{\pi} \\
p^1 &: \ u_1(x,t) = \frac{4\sqrt{t}}{\pi^{3/2}} \\
p^2 &: \ u_2(x,t) = \frac{2t}{\pi} \\
p^3 &: \ u_3(x,t) = \frac{t^2}{\pi} \\
&\vdots
\end{aligned}$$

As shown above, the sample solution for Example 4.1 with $\alpha = 0.5$ by using HPTM has been achieved. We also show our numerical result graphically in Figure 1 for the comparison of approximation between different values of fractional order.

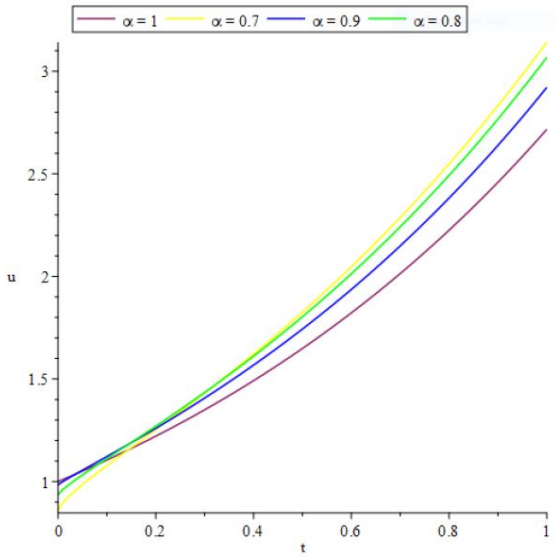


Figure 1: Comparison of solution $u(x,t)$ at different values of α in Example 4.1

From Figure 4.1, it is observed that as α decreases, the solution $u(x,t)$ increases for the same values of x and t. This indicated that when the value of $\alpha < 1$, the equation incorporates fractional effects, which increase $u(x,t)$.

Example 4.2. Consider the following time-fractional diffusion equation

$$\frac{\partial^\alpha u(x,y,t)}{\partial t^\alpha} = \frac{\partial^2 u(x,y,t)}{\partial x^2} + \frac{\partial^2 u(x,y,t)}{\partial y^2} \tag{17}$$

with the initial condition $u(x,y,0) = \sin x \sin y$.

Applying the steps as explained in Section 2 and using $\alpha = 0.5$, we get

$$u(x,y,t) = \frac{2\sin x \sin y}{\pi} + \mathscr{L}^{-1}\{s^{-1/2}\mathscr{L}(D_x^2 u(x,t) + D_y^2 u(x,t))\}\}. \tag{18}$$

Substituting (12) into (18) and equating terms to the same order of p we get

$$\begin{aligned}
p^0 &: \ u_0(x,t) = \frac{2\sin x \sin y}{\pi} \\
p^1 &: \ u_1(x,t) = \frac{-8\sin x \sin y\sqrt{t}}{\pi^{3/2}} \\
p^2 &: \ u_2(x,t) = \frac{8t\sin x \sin y}{\pi} \\
p^3 &: \ u_3(x,t) = \frac{-64t^2\sin x \sin y}{3\pi^{3/2}} \\
&\vdots
\end{aligned}$$

The calculation above is the sample solution for Example 4.2 when $\alpha = 0.5$ by using HPTM. The comparison of approximation between different values of fractional order, α is presented graphically in Figure 2 below.

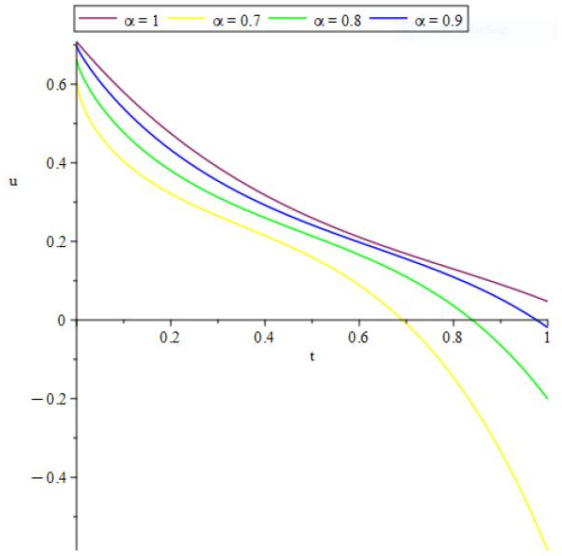


Figure 2: Comparison of solution $u(x,t)$ at different values of α in Example 4.2

From Figure 2, the solution obtained for Example 4.2 with the HPTM method appears to decrease very quickly with the increase in α at the value of $x = y = 1$. We can see a similar behaviour in the approximate solution, $u(x,y,t)$ for this example, same as in Example 4.1 when α has evolved.

Figure 3 illustrates the comparative solution between the exact solution and the approximate solution using HPTM for $\alpha = 1$.

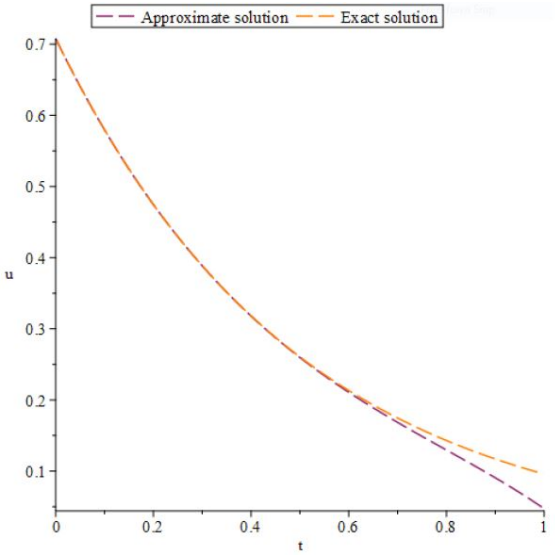


Figure 3: Comparison of exact solution and approximate solution for $\alpha = 1$

Here, we compare our solution, (18) with the exact solution, i.e. $u(x,t) = e^{-2t}\sin x \sin y$ having $\alpha = 1$ (Kumar et al., 2012). From the numerical result, we obtained a good result in comparison to the exact solution.

5. Conclusion

This study's main contribution is implementing a method to solve FDE in the Riemann-Liouville sense. In this section, the convergence of the FDE solution can be shown by using HPTM, and it proves that this method is easy to use than other methods. The efficiency of the FDE solution obtained by using HPTM with different values of α. The graph shown in Figure 1 and Figure 2 is a comparison graph of the approximate solution for different values of $\alpha = 0.7, 0.8, 0.9, 1$. When $\alpha = 1$, the equation corresponds to a standard differential equation of integer order. As $\alpha < 1$, the equation incorporates fractional derivative effects, which increase or decrease $u(x, t)$. This behaviour in fractional derivative equations shows that the order of derivative, α governs the influence of past states or memory effects on the current states of the equation.

The main goal of this study was to apply the Homotopy Perturbation Transformation Method (HPTM) to find approximate solutions for fractional differential equations (FDE). The achievement of this objective was eased by the application of the method that based on the Riemann-Liouville derivative on a specific class of FDE.

The selection of HPTM resulted from its unique feature: it balances analytical and numerical efficiency. HPTM is considered the convenient method for dealing with FDE problems. This method offers the advantage of a simpler calculation steps when compared to several other numerical and analytical techniques. The Variational Iteration Method (VIM) provides a particularly useful example for comparison. Selecting an appropriate multiplier for the mathematical problem is important in the VIM. This selection process can necessitate meaningful manual adjustment of the initial conditions. This study reveals the important potential of HPTM as an effective alternative for solving FDE, thus contributing to at least some progress in several future applications within the field of FDE.

References

Fernandez, A. and Fahad, H. M. (2022). Weighted fractional calculus: a general class of operators. *Fractal and Fractional*, 6(4):208.

Hailat, I., Nemrat, A. A., Zainuddin, Z., and Azmi, A. B. (2019). Analytical solution of ordinary fractional differential equations by modified homotopy perturbation method and laplace transform. In *Journal of Physics: Conference Series*, volume 1366, page 012037. IOP Publishing.

Javeed, S., Baleanu, D., Waheed, A., Shaukat Khan, M., and Affan, H. (2019). Analysis of homotopy perturbation method for solving fractional order differential equations. *Mathematics*, 7(1):40.

Khan, Y., Faraz, N., Kumar, S., and Yildirim, A. (2012). A coupling method of homotopy perturbation and laplace transformation for fractional models. *University" Politehnica" of Bucharest Scientific Bulletin, Series A: Applied Mathematics and Physics*, 74(1):57–68.

Kumar, S., Yildirim, A., Khan, Y., and Wei, L. (2012). A fractional model of the diffusion equation and its analytical solution using laplace transform. *Scientia Iranica*, 19(4):1117–1123.

Ma, Q., Lin, X., and Li, H. (2018). A comparative teaching of fractional calculus and integer calculus. In *2018 International Conference on Social Science and Education Reform (ICSSER 2018)*, pages 94–97. Atlantis Press.

Munoz-Pacheco, J. M., Wei, Z., Volos, C., and Sambas, A. (2023). Future challenges in the fractional-order dynamical systems: from mathematics to applications.

Oliveira, E. C. d. and Machado, J. (2014). A review of definitions for fractional derivatives and integral. *Mathematical Problems in Engineering*, 2014:1–7.

Zheng, Z., Zhao, W., and Dai, H. (2019). A new definition of fractional derivative. *International Journal of Non-Linear Mechanics*, 108:1–6.

Understanding Local Food Security Awareness: Insights from the Seremban Community

Nur Adriyana Shafie[1], Noorezatty Mohd Yusop[1,*]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban, 70300, Seremban, Malaysia*
[*]noorezatty@uitm.edu.my

Abstract

Food security has become an increasing concern in Malaysia, particularly after the COVID-19 pandemic, which exposed the vulnerability of the food supply chain. The crisis resulted in food shortages and rising prices. Considering the challenges, it is crucial to evaluate community level of understanding and preparedness in relation to food security. This study aims to assess the level of knowledge and awareness of food security among the Seremban community in Negeri Sembilan. A sample of 384 respondents was selected through convenience sampling, and data were collected via online questionnaires. The findings on awareness of the term "food security" revealed that 42.9% of respondents had never heard of it. Furthermore, awareness levels differed across demographic groups, including age, gender, education level, and household income. These demographic characteristics also showed variations in the adoption of selfsufficiency practices. The results highlight existing gaps in knowledge and behaviour among Seremban residents. The study concludes by recommending educational programmes and community-based initiatives to enhance food security awareness across all demographic groups. The findings provide valuable insights for policymakers and local leaders to design targeted interventions that strengthen community resilience and ensure long-term food security.

Keywords*:* Food security, Knowledge, Awareness level

Introduction

Food security is a fundamental issue that affects national stability, economic development, and individual well-being. In line with the United Nations' Sustainable Development Goal 2 (SDG 2), Malaysia has committed to ending hunger, achieving food security, and promoting sustainable agriculture [1]. To monitor progress globally, the Global Hunger Index (GHI) serves as a comprehensive tool for measuring and tracking hunger at the global, regional, and national levels. The GHI scores range from 0 to 100, where 0 represents no hunger and 100 indicates the most severe level of hunger. As of 2024, Malaysia has a moderate hunger level, with a GHI score of 12.7, ranking 61st out of 127 countries. This reflects a positive trend, as the country's score has improved from 15.4 in 2000 to 12.7 in 2024, indicating steady progress in reducing hunger and malnutrition. Nevertheless, ongoing challenges such as climate change, economic pressures, and rapid urbanisation continue to expose vulnerabilities in Malaysia's food security system [2].

Food security has emerged as a critical concern in Malaysia, particularly following the COVID-19 pandemic, which revealed significant vulnerabilities in the country's food supply

chain. The crisis led to food shortages and sharp increases in the prices of essential goods, disproportionately impacting rural communities [3]. With nearly 30% of Malaysia's food supply reliant on imports, global disruptions can quickly affect domestic availability and accessibility. This highlights the need for a more resilient and self-sufficient food system. This situation has shown the need to strengthen local food systems to ensure food security for all people [4]. Recent policy changes, such as the removal of chicken subsidies, have added to public concerns about food affordability, raising awareness of the importance of understanding food security at the community level [5]. Because of these challenges, studying food security is essential to assess how well communities are prepared to face future issues such as rising food prices or supply disruptions. Local awareness and active community involvement play a crucial role in promoting sustainable practices, such as home-based food production, which can reduce dependence on imports and strengthen resilience against economic shocks.

While government initiatives and community programmes such as urban farming and awareness campaigns are aimed to improve food security, there is limited evidence on the extent of knowledge and awareness among communities regarding this vital issue. Understanding the aspects is essential as knowledge and awareness greatly influence individuals' behaviours and attitudes toward food security initiatives [6]. Despite the growing attention to food security, awareness levels vary across different segments of the population. Urban residents are often more engaged in activities such as home gardening and food waste reduction; however, they frequently encounter challenges such as limited space and insufficient policy support [7]. Women, who typically manage household food-related responsibilities, tend to possess strong practical skills but may lack a comprehensive understanding of broader food system issues or national policies [8]. Addressing the gaps in knowledge and engagement is vital for Malaysia to strengthen its long-term resilience against food insecurity [9].

As food security becomes an increasingly pressing issue in Malaysia, it is crucial to understand how individuals from different communities perceive this challenge and what actions they are prepared to take in response. This study focuses on residents of Seremban, aiming to explore their knowledge, awareness, and daily practices related to food security. By identifying gaps in understanding and behaviour, the research seeks to support local initiatives in building stronger and more resilient communities.

Methodology

Study Area

This study was conducted in Seremban, the capital of Negeri Sembilan, Malaysia. Known for its socioeconomic diversity and urban growth, Seremban offers a representative setting for assessing urban community knowledge of food security.

Data Collection and Sampling

A survey was conducted to collect primary data. Data were collected through a structured, selfadministered questionnaire distributed online via Google Forms. This study employed convenience sampling, a practical and cost-effective non-probability method that selects

participants based on their availability and willingness to participate. This approach was efficient for gathering data on food security awareness among residents of Seremban, with 384 respondents from different age groups, genders, education levels, and household income groups participating in the survey. However, as a non-probability method, convenience sampling limits the ability to generalise findings to the wider population. Despite this limitation, the method is suitable for exploratory research as it provides initial insights and identifies patterns that can guide future, more comprehensive studies. The questionnaire was divided into two sections: (1) Section A collected demographic information, and (2) Section B contained questions related to knowledge and awareness of food security. Questions on awareness of food security were in the Likert-scale format. Prior to full deployment, the questionnaire was pilot tested with 30 participants to ensure clarity and reliability.

Method of Analysis

Descriptive analysis using frequency tables and bar charts was employed to summarise the respondents' demographic profiles and their level of knowledge and awareness regarding food security. This method provides a clear overview of the data by highlighting general trends and patterns. This approach enables the identification of gaps in awareness and understanding among various segments of the population.

Results and Discussion

Profiles of Respondents

Table 1 presents a descriptive analysis of the demographic profile of the 384 respondents who participated in the study.

Table 1: Demographic profiles of respondents

Variable		**N**	**Percentage**
Age Group	19-25 years	106	27.6
	26-35 years	105	27.3
	36-55 years	100	26.0
	56-60 years	73	19.0
Gender	Female	199	51.8
	Male	185	48.2
Education Level	Primary	41	10.7
	Secondary	67	17.4
	Diploma or equivalent	91	23.7
	Bachelor's degree	88	22.9
	Postgraduate or higher	49	12.8
Household Income Group	No formal education	48	12.5
	B40	154	40.1
	M40	120	31.3
	T20	110	28.6

The analysis aims to provide an overview of the demographic composition of the respondents. The age distribution is balanced, with the largest group aged 19–25 years (27.6%), followed closely by those aged 26–35 years (27.3%) and 36–55 years (26.0%). The smallest age group comprises individuals aged 56–60 years, accounting for 19.0% of the sample. In terms of gender, females slightly outnumber males, with 51.8% and 48.2% of respondents respectively. Educational attainment varies, with the highest proportion holding a diploma or equivalent (23.7%) or a bachelor's degree (22.9%). Other respondents reported having secondary education (17.4%), postgraduate qualifications (12.8%), primary education (10.7%), or no formal education (12.5%). Regarding household income, the highest category is the B40 income group (below RM5,245), representing 40.1% of the sample. The M40 group (RM5,256–RM11,819) constitutes 31.3%, while 28.6% belong to the T20 group (above RM11,820).

Knowledge of the Term Food Security

Understanding public awareness of food security is essential for developing effective strategies to ensure sustainable food access and community resilience. Figure 1 illustrates the respondents' level of awareness regarding the term "*food security*", revealing varying degrees of familiarity and understanding among the population.

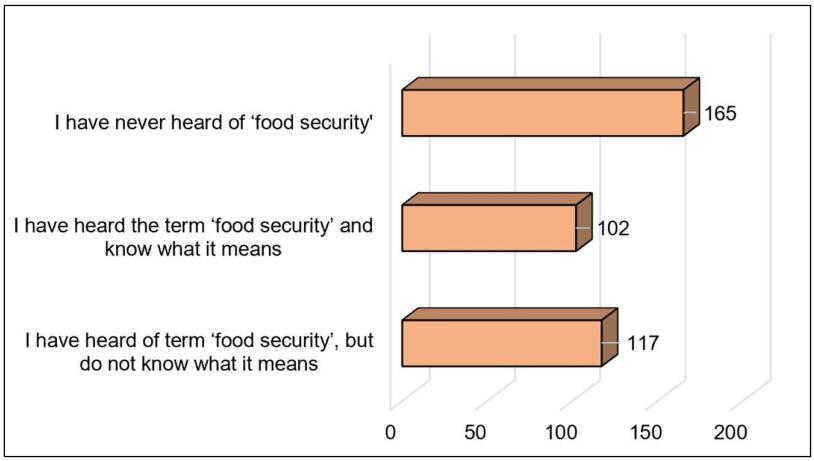


Figure 1: Knowledge of the term food security among Seremban residents

Based on Figure 1, 165 respondents (42.9%) reported that they had never heard of the term "*food security*". This indicates that a significant proportion of Seremban residents lack basic awareness of the concept, highlighting a crucial gap in public understanding. However, it is also possible that some individuals, although unfamiliar with the term itself, may already be practising food security measures, such as home gardening, food preservation, or reducing food waste. Meanwhile, 117 respondents (30.5%) stated that they had heard of the term but did not know what it meant.

This suggests that although some awareness exists, understanding remains limited, possibly due to a lack of exposure to formal education on food systems, sustainability, or relevant

government policies. In contrast, only 102 respondents (26.6%) indicated that they understood the meaning of food security, representing a relatively well-informed segment of the population. This group may be more likely to engage in food-related discussions or practices, potentially influenced by higher levels of education, occupation, or personal interest.

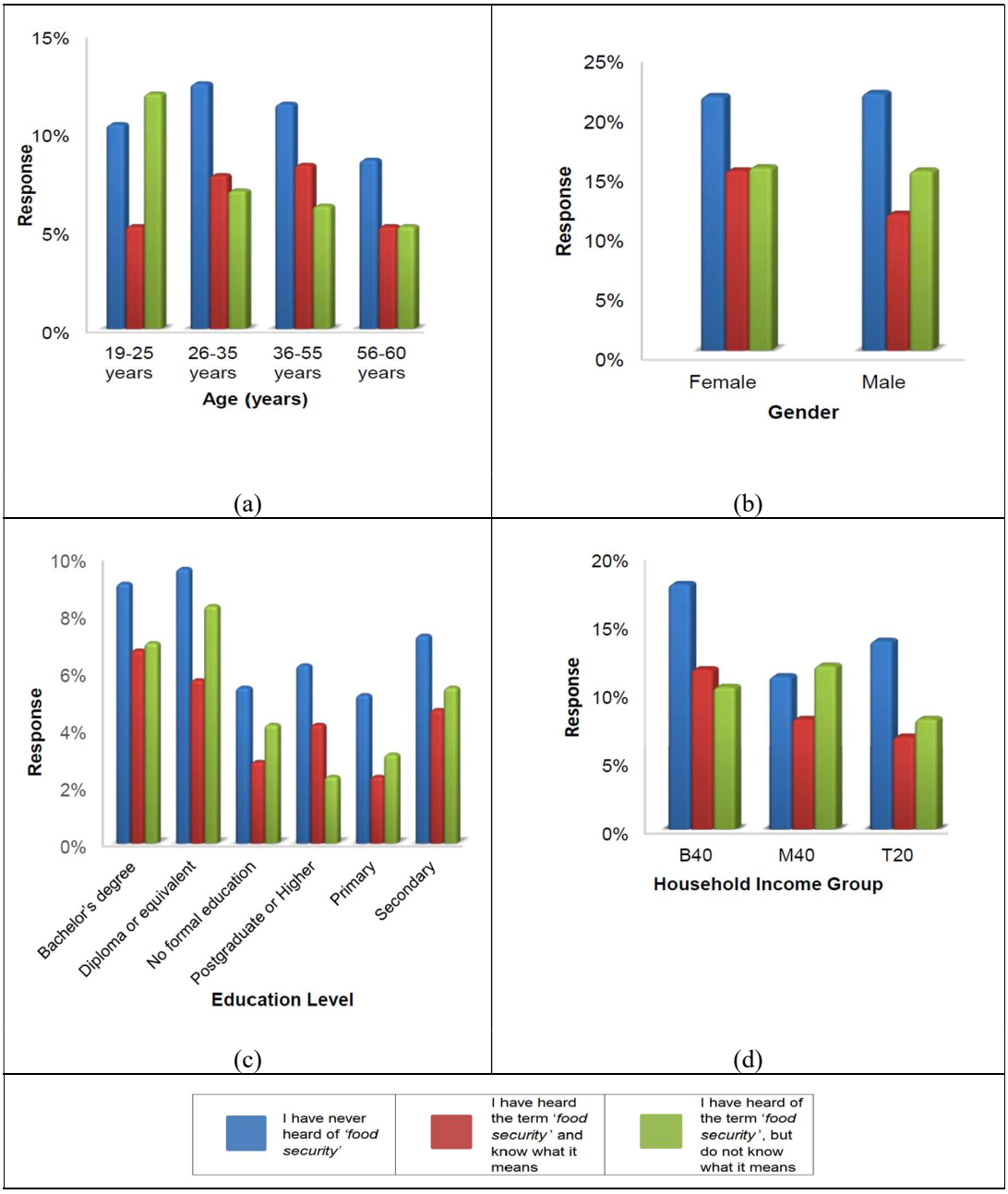


Figure 2: Knowledge of the term "food security" by demographic factor

As shown in Figure 2(a) – Figure 2(d), knowledge levels varied across different demographic groups. Respondents aged 36–55 had the highest understanding, possibly due to their exposure to current issues through education or media. In contrast, the 26–35 age group showed the lowest knowledge, indicating that younger adults may need more targeted outreach. In terms of gender, female respondents were slightly more knowledgeable than males. Furthermore, those with diplomas and bachelor's degrees showed the highest levels of understanding. However, this same group also had the highest number of respondents who had never heard of the term, showing mixed levels of awareness. The B40 (lower-income group) had the highest proportion of respondents who had never heard of food security. B40 also shows the highest understanding compared to M40 and T20, likely due to better access to practices and information. These results underline the importance of tailored outreach and education efforts that account for both economic and social conditions.

Awareness of Adoption of Self-sufficiency Practices

Food-related practices are important for food security. This section explores the awareness and practices related to food security among respondents, including food prepping practice, keeping or storing extra food and growing plants for personal use. The analysis is categorised by household income group, gender, and educational level.

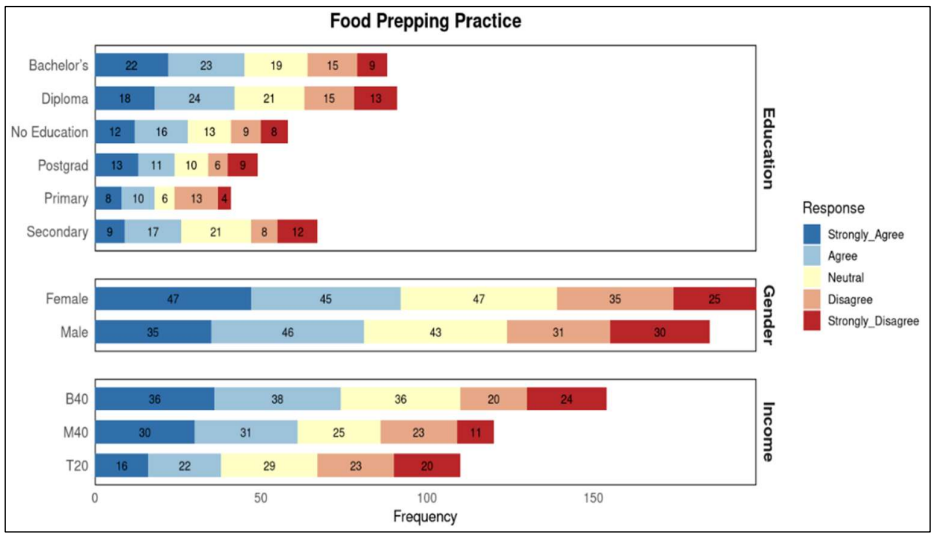


Figure 3: Awareness level of food prepping practices by demographic group

Food prepping, which involves preparing meals or ingredients in advance, is crucial for ensuring consistent access to nutritious and affordable food, especially for households with limited resources. Figure 3 shows a significant gender gap in food prepping, with more female respondents (n=92) than male respondents (n=81) agreeing with the practice. This illustrates ongoing gender norms, as more women reported practising it. According to the Theory of Planned Behavior (TPB), this is influenced by subjective norms (women expected to cook), positive attitudes (savings, convenience), and perceived control (food literacy, planning skills)

[10]. Among income levels, low-income respondents in the B40 category (n=74) demonstrate the highest awareness of food prepping, likely using it as a cost-saving strategy. Additionally, individuals with a diploma (n=42) and degree qualifications (n=45) report the highest participation, suggesting that education influences awareness of prepping practices.

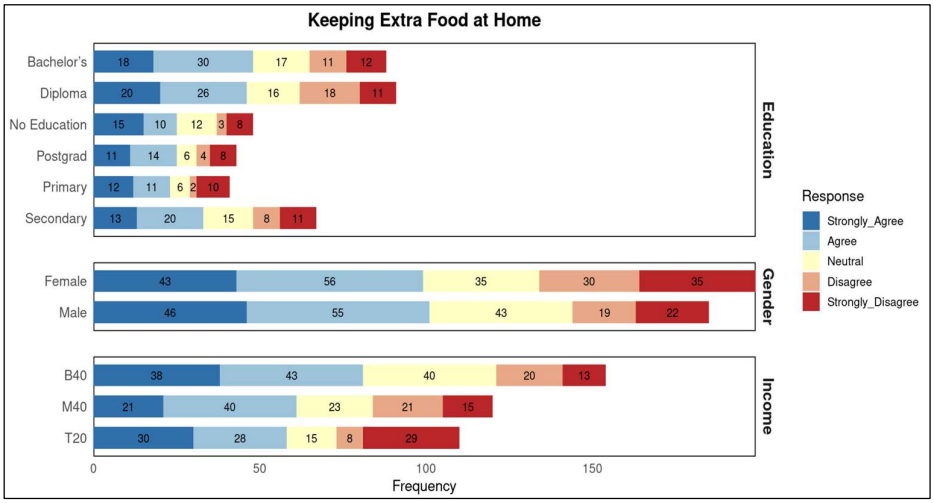


Figure 4: Awareness level of keeping extra food by demographic group

Based on the analysis presented in Figure 4, the prevalence of home food storage was relatively comparable between male and female respondents, with 101 males and 99 females indicating participation. Although this discrepancy is minimal, it suggests that both genders are actively involved in food preparedness. In terms of income group, individuals belonging to the B40 income group (n=81) demonstrated the highest level of engagement in food storage. This trend may be attributed to their enhanced capacity to purchase goods in bulk and their inclination to prepare for uncertain circumstances. Furthermore, educational attainment played a significant role, as respondents holding degrees or diplomas were more likely to engage in food storage practices. This observation indicates that individuals with higher levels of education may possess a greater awareness of the importance of food planning and preparedness for emergencies. Education is important in improving people's knowledge and awareness of food security. Research shows that individuals with higher levels of education are more likely to understand how food systems work and are more willing to follow sustainable practices [11]. They are often better informed about nutrition, environmental issues, and the importance of responsible food choices.

Figure 5 indicates that 72 female respondents grew plants for personal use compared to 65 males, suggesting greater involvement of women in home gardening. The B40 income group showed the highest agreement, with 55 individuals likely to grow their own plants. Respondents with diplomas (n=28), bachelor's degrees (n=27), and secondary education (n=27) expressed the highest agreement that growing plants was a practice for food security. This implies that higher education levels may lead to more sustainable habits such as home gardening, as these individuals are generally better informed about its benefits.

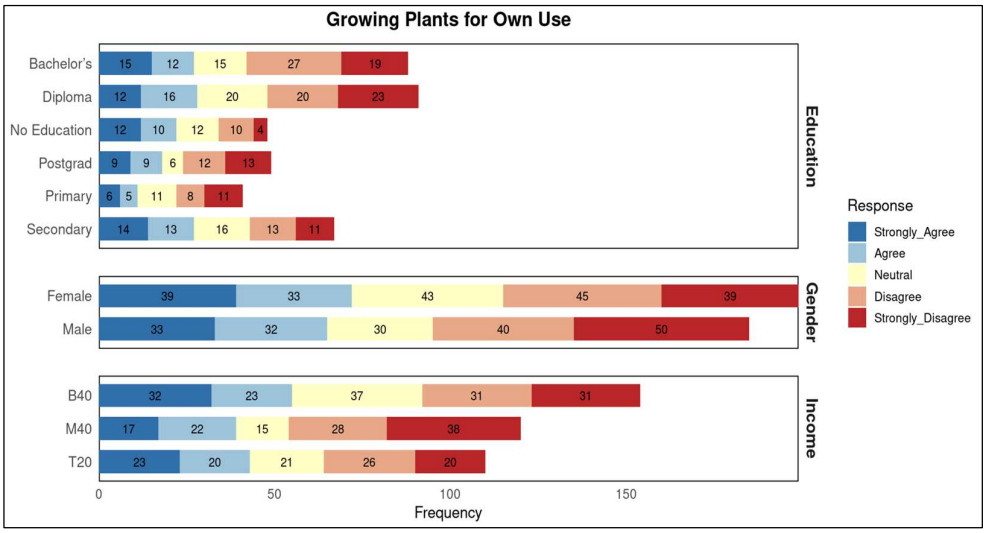


Figure 5: Awareness level of growing plants practices by demographic group

Conclusion

This study uncovers significant gaps in food security knowledge among residents of Seremban, with 42.9% unaware of the term "food security" and only 26.6% demonstrating a clear understanding. Despite this limited awareness, many individuals, particularly women, those from the B40 income group, and respondents with higher educational attainment actively practice forms of food self-sufficiency, such as food prepping, home gardening, and food keeping. The findings suggest that while conceptual understanding may be lacking, practical engagement with food security behaviours exists and is influenced by key demographic factors, such as income level, education, and gender.

In light of the findings, targeted initiatives are essential to enhance food security literacy and promote sustainable practices. Public education campaigns should be developed to introduce the concept of food security in a relatable and accessible manner, using community workshops, schools, and local media. Support for low-income households through microgrants or starter kits can encourage food resilience activities at home. Furthermore, policies that recognise and incentivise sustainable household practices, particularly those led by women, should be integrated into local governance frameworks. The efforts will contribute to strengthening community resilience and advancing Malaysia's commitment to Sustainable Development Goal 2. Future research could expand on the findings through quantitative approaches that test relationships between demographic factors and food security practices, or through longitudinal studies that track changes in awareness and behaviour over time.

Acknowledgments

Special thanks to the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Seremban Campus, for academic support and guidance.

References

[1] United Nations. (2023). The Sustainable Development Goals Report 2023. New York: United Nations. https://unstats.un.org/sdgs/report/2023/

[2] Luqfi Nulhakim Mohd Azman, A., Aqilah Abdullah, N., Khadijah Ninam Shah, S., & Puteh, F. (2023). Strategic analysis of climate change and food security in Malaysia. *Journal of Administrative Science*, 20(1), 51–77.

[3] Shah, A. U. M., Safri, S. N. A., Thevadas, R., Noordin, N. K., Abd Rahman, A., Sekawi, Z., ... & Sultan, M. T. H. (2020). COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government. *International Journal of Infectious Diseases*, 97, pp. 108–116.

[4] Ministry of Agricultural and Food Security. (2021). National Agrofood Policy 2021–2030 (NAP 2.0). https://www.mafi.gov.my

[5] Malay Mail. (2023b, October 30). Chicken subsidy removal: More socio-economic, welfare initiatives from savings. Malay Mail. https://www.malaymail.com/news/malaysia/2023/10/30/chicken-subsidy-removal-more-socio-economic-welfare-initiatives-from-savings-says-agriculture-and-food-security-ministry/99317

[6] Sklar, E., Chodur, G. M., Kemp, L., Fetter, D. S., & Scherr, R. E. (2025). Food acquisition coping strategies vary based on food security among university students. *Current Developments in Nutrition*, 9(1).

[7] Otieno, B. A., & Namiripo, D. B. (2019). The multifaceted impact of urban gardening for enhancing environment, health, and community. *Advances in Herbal Research*, 2(1), 1-5.

[8] Ya, R. M., Nor, N. M., Jaafar, N. H., Sidek, S., Rahman, J. A., Sulaiman, N., & Bakar, W. A. M. A. (2023). Validation of the malay version of food insecurity experience scale (M-FIES) using rasch analysis. *Malaysian Journal of Nutrition*, 29(3), 529-537.

[9] Sulaiman, N., Yeatman, H., Russell, J., & Law, L. S. (2021). A food insecurity systematic review: experience from Malaysia. *Nutrients*, 13(3), 945.

[10] Mclaughlin, C., Tarasuk, V., & Kreiger, N. (2003). An examination of at-home food preparation activity among low-income, food-insecure women. *Journal of The American Dietetic Association*, 103(11), 1506–1512.

[11] Abu Bakar, W. A. M., Ismail, S., Sidek, S., & Rahman, R. A. (2019). Prevalence and factors affecting food insecurity among university students in Pahang, Malaysia. *Malaysian Journal of Nutrition,* 25(1), 59–68.

ANALYZING FOOD SECURITY OF THE HOUSEHOLDS: A CASE STUDY IN MALAYSIA USING A LOGISTIC REGRESSION APPROACH

Julia Johari[1], Jaida Najihah Jamidin[2*]

[1,2] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia*

[*]jaida5698@uitm.edu.my

Abstract

Food security is a state where all households have access to sufficient, nutritious, and safe food, physically, socially, and economically. Despite being one of the highest palm oil-producing countries in Asia, Malaysia still shows inconsistent food security. Malaysia's growing population increases food demand but also raises challenges such as food insecurity, fuelled by low income, limited nutritional knowledge, and crisis such as COVID-19. To improve food security among households, the country needs to reduce imports, making food easier to access and affordable. Therefore, this study was conducted to identify the factors that contribute to food insecurity among Malaysians. Logistic regression was handled with independent variables which are gender of household head, household size, the income of the household, access to credit, age of household head, and level of education of household head towards food security. This study was carried out on Malaysian households. The result of this study revealed that only several variables namely household size, income of the household, and level of education contributed significantly to food security in Malaysia.

Keywords*:* Food Security, Logistic Regression Model

Introduction

The growing global population increases the demand for various necessities, particularly food. However, due to limited resources and limitless human desire, not every need can be fulfilled. Food plays a vital role in our daily lives, as it is consumed by everyone on a regular basis. It offers a multitude of benefits to consumers, providing essential energy required for various activities and tasks. Without adequate nutrition from food, our bodies would lack the necessary nutrients needed for growth and overall health. Additionally, a well-balanced diet supports proper breathing and enables optimal brain function. In essence, food is indispensable for sustaining life and promoting overall well-being.

Food security is a state in which all households have access to sufficient, nutritious and safe food, physically, socially, and economically. It is also defined as "ensuring that everyone has physical and economic access to the basic food they need at all times" [1]. This is a fundamental necessity that every nation must address to guarantee the access of sufficient food to all individuals. [4] defines household food security as the physical, social, and economic stability that comes from having enough safe and nutritious food available on a consistent basis to meet dietary needs and preferences, thereby promoting a healthy and productive life [4]. Additionally,

food security ensures that food remains affordable despite fluctuations in the country's currency. It is imperative to prioritize food security within households, particularly for individuals with limited incomes who are vulnerable to food insecurity.

In strengthening the country's food security, the National Food Security Framework (KSMN) has developed four main components, namely (1) availability which includes the provision of food stocks, import production, and food aid; (2) access which refers to infrastructure, logistics, and consumers' ability to afford goods; (4) utilization which involves processing and preparation methods that comply with existing standards such as HALAL; and (4) stability which encompasses sustainability and the capacity to withstand crises such as natural disasters, climate change and agricultural diseases ore pests. The main component is in line with the latest concepts recommended by the Food and Agriculture Organization of the United Nations (FAO). The annual rise in population is one of the primary reasons why food insecurity has become a worldwide concern.

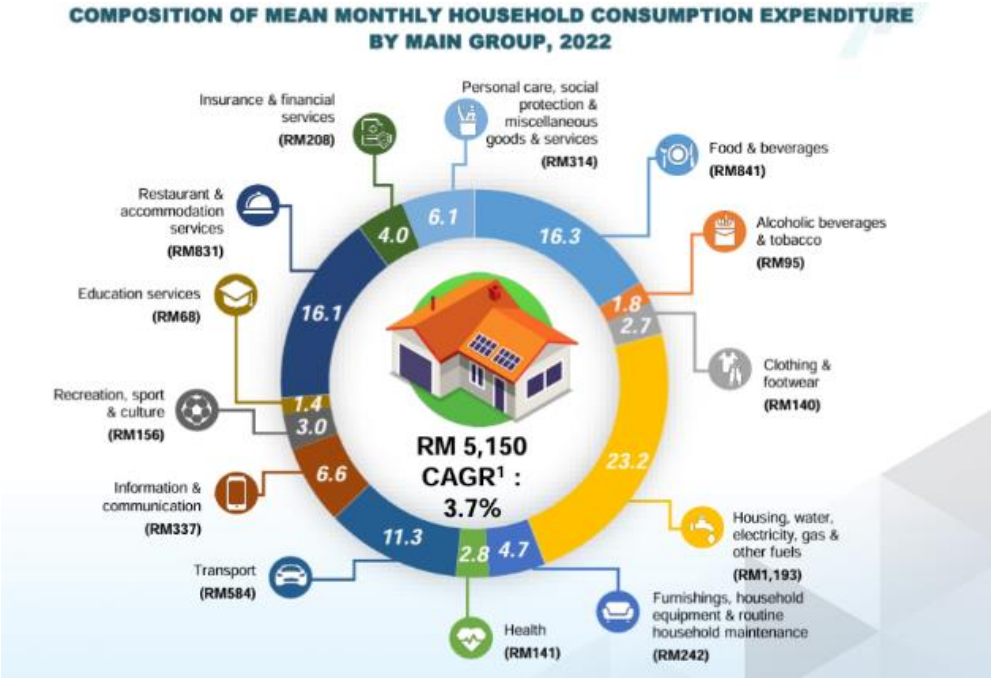


Figure 1: Ranking of Food Security Index based on Countries.

As indicated in Figure 1, Malaysian households spent an average of RM841 per month on food, which represented a significant portion of their expenses [2]. While higher-income families could be secured with this amount, this high burden might influence low-income households in terms of their inability to meet other essential needs. Hence, it pushes them toward food insecurity. These economic challenges could grow further if inflation causes the rise of food prices, particularly among the most vulnerable people in society. As a result, specific measures such as subsidies, food assistance programmes, and price restrictions may be required to ensure that all households can afford and have access to healthy foods, which are the basis of overall health and economic stability.

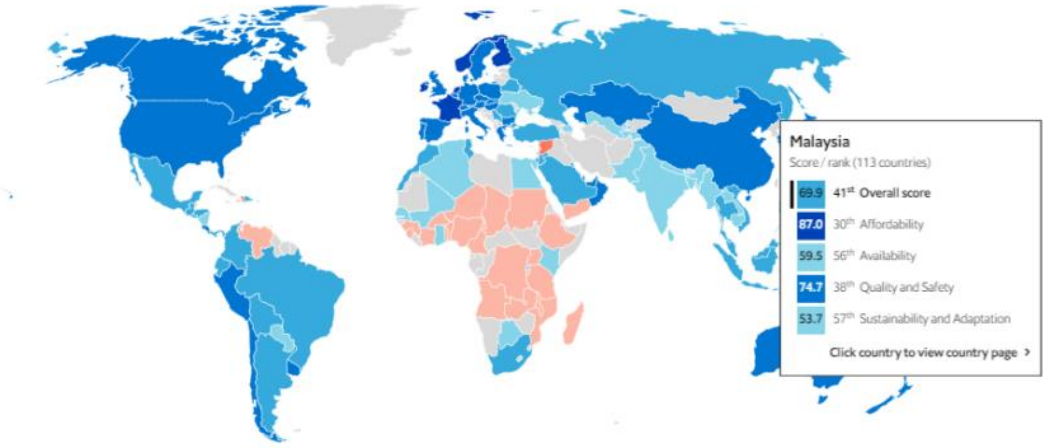


Figure 2: Ranking of Food Security Index based on Countries

Furthermore, based on Figure 2, in 2022 Malaysia is ranked 41 out of 113 countries, which is a drop of seven places compared to the position in 2021 [3]. The situation raises concerns about the security of the availability of food for Malaysian households and how they manage to keep their household food safe. This study therefore seeks to figure out the major factors that contribute to food insecurity.

Methodology

Description of Data

The data used in this study was collected from a survey of 479 respondents in Malaysia. The dataset includes variables such as household income (continuous, RM), household size (continuous, person), gender of the household head (categorical, male or female), age of household head (continuous, years), and credit card usage (categorical, yes or no). Level of education was coded as "SPM," "Diploma," Degree," and "Postgraduate". For food security status, 1 if the household food is secure and 0 if the household food is not secure. Besides, the study was conducted from October 2024 to August 2025. The study focused on the Malaysian population aged 18 and above.

Logistic Regression Analysis

Logistic regression analysis was a type of binary classification regression in which the dependent variable had only two possible outcomes. This analysis was used to define variables and determine the relationship between a dependent binary variable and its independent variables. Furthermore, this analysis was distinct from linear regression, which produced a continuous output.

The general form of logistic regression model is:

$$logit\ (P) = log\left[\frac{P(X)}{1 - P(X)}\right]$$

$$= B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K + \varepsilon$$

The above model can be rearranged to simplify it, and P can be calculated using the following formula:

$$P = \frac{e(B_0 + B_1X_1 + B_2X_2 + \cdots + B_6X_6 + \varepsilon)}{1 + e(B_0 + B_1X_1 + B_2X_2 + \cdots + B_6X_6 + \varepsilon)}$$

where,

P = Probability of household food is secure
B_0 = Constant of the equation
$B_1, \ldots, B_k$ = The coefficient of independent variable
X_1= Household Income
X_2= Household Size
X_3= Gender
X_4= Age
X_5= Credit Card Usage
X_6= Level of Education

Five factors were considered for evaluating the model. These criteria were used to assess the logistic regression model. Firstly, Omnibus Test was applied to assess if any regression coefficients significantly differ from all other coefficients. This test is commonly applied in ANOVA and multiple linear regression models. Regarding the hypotheses, if the p-value is less than the alpha value which is set to be 0.05, there was a relationship between the variables. Conversely, if the p-value exceeds 0.05, it indicates there is no relationship between the variables.

Secondly, to test the goodness of fit of a logistic regression model, the Hosmer and Lemeshow test can be used. This statistical test assesses whether the data fits the logit model well. For the regression model to be considered a good fit, the p-value (or significance value) must be greater than 0.05. Conversely, if the p-value is less than 0.05, the regression model is not well fitted. A significant result indicates that the model did not fit the data well, whereas a non-significant result indicates that it does.

Thirdly, predictive efficiency is used for model's ability to accurately predict outcomes or make predictions on new data or unseen data. It measures how well the model performs in terms of predictive accuracy and generalization. A model is considered to have good predictive efficiency if the aggregate percentage was 60% or higher. If it is less than 60%, the model is not considered to have good predictive efficiency. In addition to the overall percentage, sensitivity, specificity, and accuracy can also be used to assess the predictive efficiency of a model. To

calculate these metrics, some terms need to be defined first. True Positive (TP) is when the model accurately predicts the positive class. True Negative (TN) is when the model accurately predicts the negative class. A False Positive (FP) occurred when the model predicts the positive class incorrectly. False Negative (FN) occurs when the model incorrectly predicted a negative class. Sensitivity indicates the model's ability to predict a positive outcome correctly, such as correctly identifying households that were food-secure (coded as '1'). Specificity, on the other hand, measures the model's ability to predict a negative outcome correctly, such as identifying households that are food-insecure (coded as '0'). Accuracy represents the overall correctness of the model and indicates the proportion of correctly classified instances (both positive and negative) out of the total number of instances.

Next, Cox and Snell $R^2$and Nagelkerke R^2were methods that display the amount of variation of the dependent variable that can be explained by independent variables in the current model. These values range from 0 to 1, representing the proportion of variability in the dependent variable that is accounted for by the independent variables.

Lastlyt, another popular method for testing variable significance in logistic regression is the Wald statistic. It evaluates each regression coefficient separately in order to determine the significance of an independent, regressor, or predictor variable. When the p-value is less than the 0.05, the variable is considered a significant predictor. In contrast, if the p-value exceeds the 0.05, the variable was considered not significant. Thus, significant predictors were included in the regression model, while non-significant predictors are excluded.

Results and Discussion

Criteria for Model Evaluation

The logistic regression model was evaluated by using five model evaluation including Omnibus Test, Hosmer and Lemeshow Test, Predictive Efficiency Model, and Cox and Snell $R^2$and Nagelkerke R^2.

Table 1: The Results of Criteria for Model Evaluation

The Criteria for Model Evaluation	Full Model
Omnibus Test	0.0000
Hosmer and Lemeshow Test	0.0560
Predicitve Efficiency Model (%)	76.10
Cox and Snell R^2	0.2720
Nagelkerke R^2	0.3750

Based on Table 1, the results highlight the key criteria for the model evaluation. According to the Omnibus Test results, the p-value is 0.000, which is less than the 5% level of significance. This provides sufficient evidence to support the conclusion that improved food security predictions can be made using the information from both independent variables. The Hosmer-

Lemeshow Test results show a p-value of 0.0560, which is greater than the 5% significance level. This suggests that there is insufficient evidence to reject the null hypothesis that the observed and predicted occurrence rates differ significantly. In other words, the model fits the data well. In terms of predictive efficiency, the model achieves an overall predictive accuracy of 76.10%, which exceeds 60%. Therefore, it can be concluded that the model has strong overall predictive efficiency in determining food security status. The results for the Cox and Snell R^2 value are 0.2720, while the Nagelkerke R^2 value is 0.3750. These values indicate that the variation in food security status explained by the model ranges from 27.20% to 37.50%. The model explains food security variations based on factors such as household size, household income, and the education level of the household head (SPM). Additionally, other variables, including gender (female), credit card usage (yes), age of the household head, and higher education levels of the household head (diploma and degree), may further influence the model's predictive value.

Fitting Logistic Regression Model

The purpose of this study was to determine the best logistic regression model for the food security of households in Malaysia as well as to examine the factors such as gender of household head, household size, household income, credit card usage, age of the household head, level of education (SPM, Diploma, Degree, Postgraduate) that significantly impact this food security status of the households in Malaysia. To simplify and make the conclusion easier to understand, Table 2 was created.

Table 2: Estimated Coefficient and Wald Statistics

Variable	Estimated Coefficient	P-Value for Wald Statistics
Household Income	0.4360	0.0000
Household Size	0.2020	0.0060
Gender of the Household Head	0.3730	0.1200
Age of the Household Head	0.0120	0.2520
Credit Card Usage	0.4640	0.1080
Level of Education (SPM)	-1.1260	0.0210
Level of Education (Diploma)	-0.4300	0.3770
Level of Education (Degree)	-0.6870	0.1670
Constant	-2.0890	0.0670

Table 2 shows that the variables gender of the household head, age of the household head, credit card usage, level of education (Diploma), and level of education (Degree) are five variables that do not statistically significantly explain the food security of the households in Malaysia. This is because, compared to the other variable, their p-value is greater than 0.05. Out of the six factors mentioned in the previous study's literature review, it can be concluded that the three variables household income, household size, and level of education (SPM) are the main factors that

influence food security in households in the respondent's view. The logistic model equation used for the factors that influence food security is as follows.

Z = - 2.0890 + 0.4360 Household Income + 0.2020 Household Size + 0.3730 Gender of the Household Head(female) + 0.0120 Age of the Household Head + 0.4640 Credit Card Usage (Yes) – 1.1260 Level of Education (SPM) – 0.4300 Level of Education (Diploma) – 0.6870 Level of Education (Degree).

Conclusion

To sum up, the logistic regression analysis was successfully conducted to achieve the objective of identifying the factors which are gender of household head, household size, the income of the household, access to credit, age of household head, and level of education of household head that contributed to food insecurity in the household among respondents. Six assumptions were checked before completing the logistic regression analysis, and all of the assumptions were met. Among all of the factors tested, there were only three significant factors, which were household size, income of the household head, and level of education (SPM).

As a suggestion to future research, identifying additional factors that contribute to household food insecurity in Malaysia remains a valuable area of study. As food security remains a crucial issue nowadays, further study can highlight its significance in ensuring that all citizens have consistent access to adequate, safe, and nutritious food. Furthermore, promoting public awareness through campaigns and educational initiatives can increase the impact of study findings by encouraging shared knowledge and effort to attain national food security.

Conducting long-term analysis is another recommendation for future studies. Weekly, monthly, or yearly data analysis can provide greater insights into food security trends and patterns, allowing households to be classified as food secure or insecure more accurately. Such longitudinal research would provide an understanding of how food security changes in response to diverse socioeconomic and environmental factors. In addition, future research must emphasize the use of larger sample sizes. A larger dataset will produce results that are more relevant and representative of Malaysia's diverse population. Small sample sizes may not capture the full range of factors influencing food security, reducing the robustness of statistical models.

Finally, the adoption of sophisticated robust statistical approaches is strongly recommended. In this investigation, a large number of observations were removed due to the existence of outliers, which influenced the final model. Implementing robust statistical procedures, such as those that are resistant to outliers, may assist in reducing the problem and preserving useful data. For example, techniques such as Mahalanobis distance or other outlier detection methods could be used to successfully identify and manage outliers, ensuring that critical independent variables are kept while logistic regression assumptions remain successfully met. Finally, using probability sampling techniques is advisable to minimize bias.

References

[1] Battersby, J. (2012). Beyond the food desert: Finding ways to speak about urban food security in south africa. *Geografiska Annaler: Series B, Human Geography*, 94(2), 141–159.

[2] Chee, J.-e. (2024, March). What is the average household expenditure in malaysia? [Accessed: 2024-10-08]. https://dollarsandsense.my/what-is-the-average-household-expenditure-in-malaysia/

[3] Global Food Security Index (GFSI). (2022). https://impact.economist.com/sustainability

[4] World Food Summit Declaration. (1996). *Rome declaration on world food security*.

A DECISION-MAKING FRAMEWORK FOR SMALL RESTAURANT LOCATION SELECTION USING THE ANALYTIC HIERARCHY PROCESS

Anis Amira Asri[1], Nurul Atikah Amran[2], Nursyamimi Mahzan[3], Siti Sarah Januri[4,*]
[1,2,3,4]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan*
[*]sarahjanuri@uitm.edu.my

Abstract

Small restaurant businesses often encounter challenges in achieving sustainability and success. Among these, poor location selection ranks as the leading cause of failure, given that location is a strategic decision that anchors the business. The objectives of the study are to identify the most influential factor for location selection, and to determine the most suitable location type for small restaurants. Surveys were conducted with small restaurant owners from three distinct location types such as industrial, residential, and commercial. The analysis revealed that location characteristics were the most important factor in industrial areas, cost was most significant in residential areas, and demographic attributes were most critical in commercial zones. Furthermore, Decision Maker 1 and Decision Maker 3 ranked commercial zones as the most favorable for opening a small restaurant, whereas Decision Maker 2 favored residential zones. These insights provide valuable implications for entrepreneurs and urban planners involved in small restaurant development.

Keywords: decision-making, location selection, Analytic Hierarchy Process

Introduction

With the growing number of restaurants, the industry has become increasingly competitive. Among these, the choice of location stands out as a decisive factor which often determines whether a restaurant thrives or fails in its early years. A study found that young companies have a higher probability of failing ranging between 64% and 96% as early as four years before the failure [1]. The motivation for this study comes from this critical challenge of selecting a strategically sound location that aligns with market demands and customer accessibility.

Previous studies have consistently emphasized the importance of location in establishing and maintaining a restaurant brand. [2] noted that a well-situated and visible restaurant is a key contributor to brand success, especially when located in densely populated areas. Similarly, [3] and [4] highlighted that an optimal site not only attracts a steady customer base, but also improves a business's competitiveness and profitability. Location decisions are inherently strategic as they anchor a business in space and directly influence its visibility, accessibility, and long-term viability.

[5] reported that a large proportion of restaurant failures occur in the first two years of operation. This underscores the need for a robust decision-making process when se-

lecting a location. For small restaurants with limited capital and operational bandwidth, an unchosen site can be irreversible and financially devastating. Therefore, this study is timely and significant and aims to deepen the understanding of how key location factors influence success in small-scale restaurant ventures.

Specifically, the study employs the Analytic Hierarchy Process (AHP) to evaluate the importance of various factors such as cost, demographic characteristics, competition, physical features, and visibility that influence restaurant location decisions. The research focuses on three primary types of location, industrial, residential, and commercial areas, with input collected from experienced small restaurant owners in Johor and Negeri Sembilan. The objectives of this study are to identify the most influential factor for each type of location and to determine the most favorable location for small restaurant businesses based on expert evaluations.

Research Methodology

This study adopts a decision-analytic approach to understand and prioritize factors that influence the selection of small restaurant locations. To achieve this, the Analytic Hierarchy Process (AHP) was applied as a Multi-Criteria Decision Making Tool (MCDM). Given the focus of the niche on experienced small restaurant owners, a purpose-sampling was used. This method ensures that only respondents with relevant industry knowledge and experience participated in the study.

There is no general rule to determine the proper sample size for AHP, however, the selection of respondents and their understanding of the survey is essential [6]. Decision makers are selected purposefully based on their experience and knowledge of this study issue. The owners are selected based on the three different types of location which are Industrial, Residential and Commercial.

Data were collected using a structured questionnaire, adapted from the validated instrument developed by [7]. The questionnaire was divided into several pairwise comparison matrices, where decision makers compared the relative importance of five key criteria: Location Characteristics, Demography, Cost, Physical Features and Competition

Each criterion was evaluated on three potential location types, industrial, residential, and commercial. The responses were captured using the Saaty scale 1–5, ranging from equal importance (1) to extreme importance (5). To improve response rates and reduce bias, data collection was carried out using a mixed-mode approach, combining online surveys and follow-up telephone interviews [8].

Analytical Hierarchy Process (AHP) Steps

The AHP methodology was implemented as follows:
Step 1: Develop Hierarchical Structure
A hierarchy was constructed with the main goal (optimal restaurant location) at the top level, followed by the five evaluation criteria and the three alternative locations at the bottom level.

Step 2: Pairwise Comparison Matrices

Decision Makers were asked to perform pairwise comparisons of criteria and location alternatives based on their subjective judgment and practical experience.

Step 3: Normalize the Matrices

Each matrix was normalized by dividing the value of each cell by the sum of its respective column, allowing for comparison across different scales.

Step 4: Calculate the Priority Vectors

The average of each row in the normalized matrix was computed to generate the priority vector, representing the relative importance of each criterion.

Step 5: Evaluate Consistency

To ensure judgment reliability, the Consistency Index (CI) and Consistency Ratio (CR) were calculated. CR values 0.10 were considered acceptable (Saaty, 1990).

Step 6: Compute Overall Priorities

Final location rankings were obtained by aggregating the criteria weights and multiplying them by the corresponding priority scores of each location alternative.

Step 7: Interpret and Compare Results

The overall scores were then used to determine the most favorable restaurant location according to each decision maker, offering a comparative perspective across different urban zones.

Results and discussion

Demographic Profile of Decision Makers

Table 1 shows the three decision makers participated in this study, each representing a distinct restaurant location types where industrial area as DM1, residential area as DM2, and commercial area as DM3. All participants were experienced female restaurant owners aged between 36 and over 56 years. Their businesses had been in operation for at least five years, with most reporting monthly sales ranging from RM10,000 to over RM20,000 and employing between 5 to 10 full-time staff. These demographic characteristics align with the SME criteria, validating their suitability as knowledgeable evaluators in the location decision-making process.

Table 1: Demographic Profile of Decision Makers

Demographic	DM 1	DM 2	DM 3
Gender	Female	Female	Female
Age group	36 - 45 years old	$>$ 56 years old	46 - 55 years old
Restaurant Operating	$>$ 15 years	5 - 10 years	$>$ 15 years
Monthly Sales	RM10K - RM15K	RM10K - RM15K	$>$ RM20K
Number of employees	5-10	5-10	5-10

Priority Factors Across Location Types

The AHP analysis allowed the identification of the most influential factors for each decision maker based on calculated priority vectors. Table 2 shows the priority vector for

each criterion where Location Characteristics (F1), Demography (F2), Cost (F3), Physical Features (F4) and Competition (F5). DM1 (Industrial area) assigned the highest weight to Location Characteristics (priority score: 0.4363). This is expected in industrial zones, where accessibility, visibility from main roads, and logistic efficiency are crucial to attract transient workers and ensure smooth supply chain operations. DM2 (Residential area) ranked Physical Features (0.3262) as most important. For restaurants in residential neighborhoods, features such as parking availability, space layout, and proximity to households can significantly influence customer convenience and retention. DM3 (Commercial area) emphasized Demography (0.4065) as the top criterion. Commercial zones typically benefit from high foot traffic and diverse customer profiles, making it essential to understand population density, income levels, and consumption behavior. These findings reflect real-world expectations and reinforce existing literature that location-related priorities vary depending on the business environment [9, 10].

Table 2: The Most Important Factor According to Each Decision Maker

Decision maker	F1	F2	F3	F4	F5	Most important factor
DM1	0.4363	0.2125	0.1232	0.1681	0.06	F1
DM2	0.1730	0.2101	0.3262	0.1159	0.1748	F3
DM3	0.2466	0.4065	0.1840	0.1151	0.0478	F2

Best Location Alternative According to Each Decision Maker

The final rankings of the three location alternatives which Industrial (L1), Residential (L2), and Commercial (L3) were derived by calculating the overall priority vector, which aggregates the weighted contributions of each factor.

Table 3: Rank Comparison for Every Alternative

Decision maker	L1	L2	L3	Rank
DM1	0.2336	0.1301	0.6363	L3, L1, L2
DM2	0.3818	0.4612	0.1569	L2, L1, L3
DM3	0.1498	0.3711	0.4891	L3, L2, L1

Table 3 shows that DM1 ranked Commercial areas (0.6363) as the most favorable, despite operating in an industrial zone. This suggests a forward-looking perception that commercial areas offer better growth potential due to higher visibility and customer base. DM2, who operates in a residential area, ranked Residential zones highest (0.4612). Her judgment aligns with her practical experience and reflects satisfaction with the residential market's cost structure and customer consistency. DM3 also favored Commercial locations (0.4891), consistent with her operational environment. This affirms the commercial zone's capacity to support customer-driven businesses, particularly in food services.

References

[1] Abdullah, N. A. H., Ahmad, A. H., Zainudin, N., Rus, R. M. (2016). Modelling small and medium-sized enterprises' failure in Malaysia. *International Journal of Entrepreneurship and Small Business, 28(1), 101–116.*

[2] Hanaysha, J., et al. (2016). Restaurant location and price fairness as key determinants of brand equity: A study on fast food restaurant industry. *Business and Economic research, 6(1), 310–323.*

[3] Craig, C. S. (1984). Models of the retail location process: a review. *Journal of Retailing, 60(1), 5–36.*

[4] Durvasula, S., Sharma, S., Andrews, J. C. (1992). Storeloc: A retail store location model based on managerial. *Journal of Retailing, 68(4), 420.*

[5] Healy, J., Mac Con Iomaire, M. (2019). Calculating restaurant failure rates using longitudinal census data. *Journal of Culinary Science Technology, 17(4), 350– 372.*

[6] Yoon, S.-H., Park, J.-W. (2015). A study of the competitiveness of airline cargo services departing from korea: Focusing on the main export routes. *Journal of Air Transport Management, 42, 232–238.*

[7] Hsu, P.-F., Lin, F.-L. (2013). Developing a decision model for brand naming using delphi method and analytic hierarchy process. *Asia Pacific Journal of Marketing and Logistics.*

[8] Guinal´ıu, M., de Rada, V. D. (2020). Combining sources of information to increase survey response rates. *Spanish Journal of Marketing-ESIC.*

[9] Wibisono, Y., Marella, S. (2020). A decision making model for selection of cafe location: ´ An AHP approach. *In Journal of physics: Conference series (Vol. 1477, p. 052030).*

[10] Yang, Y., Roehl, W. S., Huang, J.-H. (2017). Understanding and projecting the restaurantscape: the influence of neighborhood sociodemographic characteristics on restaurant location. *International Journal of Hospitality Management, 67, 33–45.*

Forecasting Maternal Mortality Rate in Malaysia

Nurul Asyiqin Sahrulnizam[1], Nor Akmal Md Noh [1,*]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Seremban Campus, 70300 Negeri Sembilan*
[*]norak029@uitm.edu.my

Abstract

Maternal mortality remained a significant issue in Malaysia despite the advancements in healthcare. A key factor affecting maternal mortality rates is the decline in the number of pregnancies and live births. As Malaysia faces economic challenges such as rising living costs, stagnant wages, and financial instability, more women are choosing to delay or avoid pregnancy altogether. The primary objective of this study is to examine the pattern of maternal mortality in Malaysia from the year 1946 to 2023 and to estimate the maternal mortality rates for the next three years in Malaysia using Box-Jenkins methodology. The findings of this study showed that the pattern of maternal mortality rates in Malaysia from 1946 to 2023 was seen to be exponentially decaying. The data was divided into training and test sets for ARIMA analysis to assess the accuracy of the model. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were used to evaluate the data's stationarity to identify the model. Several ARIMA models were proposed, including ARIMA (1,1,1), ARIMA (1,1,2), ARIMA (2,1,1) and ARIMA (2,1,2). Based on its capacity to produce white noise errors, its effectiveness in fitting the data, and its precision in producing a precise forecast, the ARIMA (1,1,1) model was determined to be the best model. ARIMA (1,1,1) was determined to be the best model for forecasting maternal death rates in Malaysia based on the forecast accuracy and how well the model's prediction matched the actual data. Maternal mortality rates were estimated to be 20.2, 24.6 and 21.1 in 2024, 2025 and 2026 respectively. Reducing maternal mortality requires better healthcare access, training, and public awareness. Future research should explore advanced models and higher-frequency data for improved accuracy.

Keywords*:* Maternal Mortality, Forecasting, Box-Jenkins Methodology, ARIMA model

Introduction

Maternal mortality remains a significant health issue globally, especially in developing countries like Malaysia. Maternal mortality, defined as the death of a woman during pregnancy or within 42 days after childbirth, regardless of the stage or location of the pregnancy, resulted from a cause linked to or worsened by the pregnancy or its care, rather than from an accidental or unrelated cause. In Malaysia, maternal mortality rates are measured by the number of maternal deaths per 100,000 live births [1]. Despite medical advancements, socioeconomic factors, healthcare accessibility, and quality of maternal services, it still influences maternal mortality rates [2].

The United Nations' Sustainable Development Goals (SDGs) aims to the worldwide maternal mortality ratio until it reaches less than 70 per 100,000 live births by 2030, with no country's rate surpassing double the global average [3]. As a member state, Malaysia had

committed to meeting this goal. Nonetheless, achieving this goal requires identifying high-risk groups, implementing certain policies to address issues and consistent monitoring of patterns in maternal mortality. While Malaysia has shown progress in reducing these rates, challenges persist due to economic pressures, limited rural healthcare access, and disparities in maternal care [4]. Declining fertility rates, influenced by economic uncertainty and improved access to family planning, have contributed to fewer pregnancies and reduced maternal deaths [5].

The decrease in pregnancy and birth rates affected maternal mortality statistics, but it also brought broader concerns about reproductive health and demographic patterns in the country. Although fewer pregnancies meant a reduction in pregnancy-related health complications, there was growing concern that a sustained decline in birth rates could have significant social and economic impacts over time [6]. However, this demographic shift raises concerns about an aging population and economic impacts, including labor shortages and increased pressure on public resources [7]. Addressing maternal health care gaps and improving support for working mothers are critical steps towards encouraging family planning [8]. The main objective of the study is to forecast the maternal mortality rates in Malaysia for the years 2024 until 2026.

Methodology

Data Set

The data used in this study are the yearly maternal mortality rates in Malaysia recorded by the Department of Statistics Malaysia (DOSM) from the year 1946 to 2023. The analysis includes the year and the annual maternal mortality rates. The data on maternal death cases in Malaysia from 1946 to 2023 were split into two parts in this research. In the estimation phase, the model was trained to identify the trends and patterns using 75% of the data, which covered 1946 to 2003. The evaluation phase took place over the remaining 25%, which ran from 2004 to 2023. By contrasting the model's predictions with the actual values from this frame, this phase evaluated the model's performance.

Autoregressive Integrated Moving Average

In the Box-Jenkins methodology, stationarity is one of the key assumptions, as it ensures that the statistical properties of a time series remain constant over time. Graphical tools such as the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are useful for identifying stationarity. The ACF shows the variation of the autocorrelation with different lags, where consistently low or near-zero values across all lags indicate stationarity. The PACF highlights the correlation between two points after accounting for the effects of intermediate points. If the first PACF value is close to one and subsequent values are near zero, the series is likely non-stationary. In addition, the Augmented Dickey-Fuller (ADF) test is a statistical method used to formally test for stationarity by checking for the presence of a unit root. The null hypothesis states that the data is non-stationary. If the p-value obtained from the test is less than 0.5, the null hypothesis is rejected, indicating that the data is stationary.

Model Identification

The Box-Jenkins methodology was used to identify the best-fitting model for the data. During this process, the ACF and PACF were the key tools. If the data were not stationary, differencing was applied to make it stationary. After achieving stationarity, the ACF and PACF diagrams were analyzed for patterns. If the ACF decayed exponentially and the PACF had significant spikes, the data was suited for an Autoregressive (AR) model, denoted as AR(p), where p is the number of spikes, a Moving Average (MA) model, denoted as MA(q), was used, where q represents the number do significant spikes in the ACF. The ARIMA model is expressed in equation 1 [9].

$$Y_t = Y_{t-1} + \Phi_1 Y_{t-1} + \ldots + \Phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{1}$$

where Y_t is the maternal mortality rate from 1946 to 2023, $Y_{t-1},\ldots,Y_{t-p}$ is the lagged values of the time series, $\varepsilon_{t-1},\ldots,\varepsilon_{t-q}$ is the lagged values of the error term, while, $\Phi_1,\ldots,\Phi_p$ and $\theta_1,\ldots,\theta_q$ represent the autoregressive parameters and the moving average parameters, respectively.

Model Estimation and Validations

Box-Jenkins models were estimated using sample statistics, which needed validation to ensure accuracy in the representation of the population parameters. Common statistical measures used to select and validate the best ARIMA model included Akaikes's Information Criteria (AIC), Bayesian Information Criterion (BIC), Ljung-Box Test, Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). In general, the AIC value is used to measure the fitness of the Model. While the BIC value was designed to select a model that provided the most accurate forecasts.

$$AIC = e^{\frac{2k}{T}} \left(\frac{\sum_{t=1}^{T} e_i^2}{T} \right) \tag{2}$$

The AIC equation in equation 2, $k = p + q$ gives the number of parameters estimated in the model, where p and q refer to the number of parameters for AR and MA components respectively. While T is the number of series in the dataset.

$$BIC = T^{\frac{k}{T}} \left(\frac{\sum_{i=1}^{T} e_i^2}{T} \right) \tag{3}$$

The Bayesian Information Criteria (BIC) in equation 3, k is the total number of parameters in the model, including the constant term while T is the number of observations in the series. The modified Q statistics, developed by Box and Pierce (1970) and refined by Ljung and Box (1978), tests model adequacy. Box-Jenkins assumes residuals are uncorrelated, random, and

normally distributed. If not, the model may be mis-specified, with missing or irrelevant parameters. One example of such a test was the Ljung-Box Test [9]. The Null Hypothesis is residuals were white noise.

$$Q = T(T+2)\cdot\sum_{k=1}^{m}\frac{r_k^2}{T-k} \tag{4}$$

Equation (4) shows the Ljung-Box test where r is the estimated total number of parameters, T is the number of residuals and k is the time lag. In addition, m is the number of time lags to be tested while r_k is the residual autocorrelation at lag k.

Model Comparison

The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), which are provided in equations (5) and (6), were used to evaluate the accuracy of the model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}e_i^2} \tag{5}$$

$$MAPE = \frac{\sum_{i=1}^{n}\left|e_i \middle/ y_i\right|}{n}\cdot 100 \tag{6}$$

In equations (5) and (6) the e_i represents the difference between the actual and fitted value at time period i, while is the length of the evaluation part. The error values on both equations are calculated on the fitted values for a particular forecasting method.

Results and Discussion

The first step in applying the ARIMA model is to check for the stationarity of the data. Figure 2 (a) and (b) reveal that the series is not stationary, where there are decaying patterns on the ACF diagram and the ADF test gives a p-value of 0.8653. Therefore, to proceed with Box-Jenkins modeling, it is necessary to transform the series into a stationary form. After performing first-order differencing, Figure 2 (c) and (d) illustrate that the series becomes stationary, where the decaying pattern disappears and around 1 or 2 autocorrelation values exceed the significance limits. Also, the ADF test gives p-value 0.01. As a result, the series is integrated of order one, which indicates that the data becomes stationary after one differencing. The model obtained represents the general form of ARIMA (p, d, q) with d = 1, where d represents the degree of differencing required to achieve stationarity.

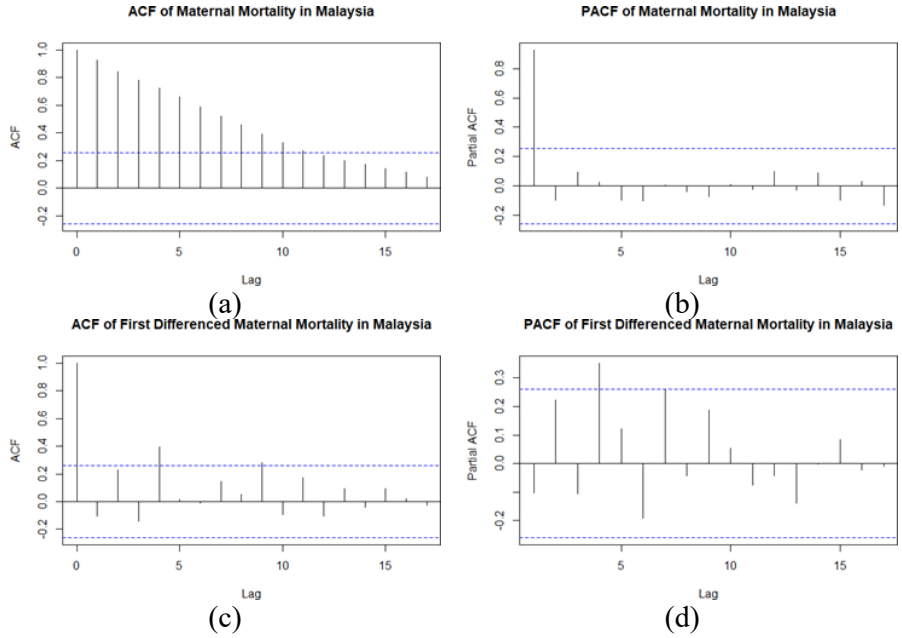


Figure 1: ACF and PACF Plots for Maternal Mortality Rates in Malaysia.

Table 1 illustrates the p-value result of the Ljung-Box Test. Based on the results, ARIMA (1,1,1), ARIMA (1,1,2), ARIMA (2,1,1), and ARIMA (2,1,2) yield p-values exceeding the threshold (0.05 value), confirming that their residuals exhibit white noise properties and validating their suitability for forecasting purpose.

Table 1: Summary of Ljung-Box Test.

Model	p-value
ARIMA (0,1,1)	0.0144
ARIMA (1,1,0)	0.0002
ARIMA (1,1,1)	0.0691
ARIMA (1,1,2)	0.0780
ARIMA (2,1,1)	0.0508
ARIMA (2,1,2)	0.1771

The best ARIMA model was assessed with the use of Akaike's Information Criteria (AIC) and Bayesian Information Criterion (BIC) to be considered for suitability in the estimation of the final parameters. Evaluating also included RMSE and MAPE values wherein preference for that model, the lowest in those values will determine the selection for the best model. As shown in Table 2 among the 6 models proposed, ARIMA (1,1,1) is identified as the most suitable model, as it exhibits the lowest BIC, Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), with values of 539.8530, 28.8517, and 0.9883, respectively. Although the AIC from ARIMA (1,1,1) is not the smallest (533.7770), the marginal difference does not outweigh the model's superior forecasting performance, making it the optimal choice.

Table 2: Summary of The Estimated ARIMA Model.

Model	AIC	BIC	RMSE	MAPE
ARIMA (1,1,1)	533.7770	539.8530	28.8517	0.9883
ARIMA (1,1,2)	533.2548	541.3562	29.2452	1.0043
ARIMA (2,1,1)	535.2686	543.3700	29.1788	0.9992
ARIMA (2,1,2)	535.9113	546.0381	29.1788	1.0009

The best model to represent the yearly maternal mortality rate in Malaysia from year 1946 to 2023 is expressed as equation (7). The same equation is used to generate the rate of maternal mortality in 2024, 2025 and 2026 as shown in Table 2.

$$Y_t = Y_{t-1} + \Phi_1 Y_{t-1} - \Phi_1 Y_{t-2} - \theta_1 \varepsilon_{t-1} + \varepsilon_t \tag{7}$$

Table 3: Summary of Forecast Value of Maternal Mortality Rate in Malaysia.

Year	2024	2025	2026
Forecast	20.2	24.6	21.1

Table 3 indicates a significant increase and decrease in the future maternal mortality rate in Malaysia per 100,000 live births. For the year 2024, there will be a significant decrease in the maternal death rate with only 20.2 mortality rate compared to 25.7 maternal mortality rate cases reported in the previous year, which is the year 2023. And in 2025, however, the trend shifted when the forecasted rate increased to 24.6 compared to the previous year. The forecast value for the maternal mortality rate in Malaysia decreases from 24.6 to 21.1 for the year 2026.

Conclusion

The maternal mortality rates in Malaysia pattern shown a distinct trend, revealing consistent changes influenced by healthcare improvements, policy implementation and socio-economic factors. This analysis highlighted areas where targeted interventions are needed to reduce maternal deaths. Additionally, the study aimed to forecast future maternal mortality rates using the Box-Jenkins Methodology, with ARIMA (1,1,1) identified as the best model based on AIC, BIC, RMSE, and MAPE values. Using this model, maternal mortality rates were forecasted for 2024 to 2026, showing a consistent range of 20 to 24 cases per 100,000 live births. To address the persistent issue of maternal mortality, the study emphasizes the need for educational campaigns to raise awareness, significant investments in healthcare infrastructure, equitable access to quality maternal care, and partnerships with international organizations to adopt comprehensive strategies. These efforts are essential for reducing maternal mortality and improving public health outcomes in Malaysia.

Acknowledgments

I would like to express my deepest gratitude to my supervisor for her guidance and support throughout this study. I am also thankful to my family and friends for their encouragement and understanding during this journey. Lastly, I extend my appreciation to everyone who contributed directly or indirectly to the completion of this project. Thank you all for your invaluable support.

References

[1] Department of Statistics Malaysia. (2024, October). Annual maternal deaths — opendosm. https://open.dosm.gov.my/data-catalogue/deaths maternal.

[2] Alvarez, J. L., Gil, R., Hern´andez, V., & Gil, A. (2009). Factors associated with maternal mortality in sub-Saharan Africa: An ecological study. BMC public health, 9, 1–8.

[3] UNICEF. (2024, May). UNICEF annual report 2023 — UNICEF. https : / / www.unicef.org / reports/unicef-annual-report-2023.

[4] Stanton, M. E., Higgs, E. S., & Koblinsky, M. (2013). Investigating financial incentives for maternal health: An introduction. Journal of Health, Population, and Nutrition, 31(4 Suppl 2), S1.

[5] Bernstein, A. (2021, May). Maternity, maternal health, and the economy during the pandemic. https : / / tcf . org / content / commentary / maternity - maternal – health - economy-pandemic/

[6] Wu, Y., Su, B., & Li, J. (2023). The impact of low fertility rates on labor demand and socioeconomic development in China. China CDC Weekly, 5(27), 599.

[7] Fang, H. (2023). The impact of population aging on financial services and economic development. Available at SSRN 4574199.

[8] Karim, Z. A., Nuruddin, N. A. M., Karim, B. A., Mohamad, M., & Ishak, I. (2023). The impact of population aging and fertility rate on economic growth in Malaysia. Economic Journal of Emerging Markets, 199–211.

[9] Lazim, M. A. (2011). Introductory business forecasting : A practical approach (3rd ed).UiTM Press.

Performance of Box Jenkins and Artificial Neural Networks Models on Air Pollution Index in Klang

Nor Akmal Md Noh [1,*], Nur Aqilah Aina Abd Latif [1]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Seremban Campus, 70300 Negeri Sembilan*
[*]norak029@uitm.edu.my

Abstract

This study focuses on air pollution levels in Klang Valley, a region frequently experiencing poor air quality due to industrial activity, haze, traffic and urbanization. The performance of Box Jenkins (ARIMA) and Artificial Neural Network (ANN) models was compared using data collected from Malaysia's Department of Environment (DOE) over five years (2019-2023). The ARIMA model analyzed Air Pollution Index (API) as the variable, while the ANN model utilized API as the output variable and six pollutants sulphur dioxide (SO_2), particulate matter (PM_{10}), ($PM_{2.5}$), ozone (O_3), nitrogen dioxide (NO_2) and carbon monoxide (CO) as input variables. Under Box Jenkins procedure, six ARIMA models were tested, with ARIMA (4,1,1) selected as the best based on AIC and BIC. As an ANN model, networks with varying hidden nodes were evaluated, and the model with five hidden nodes was identified as the best, achieving highest R^2 and lowest RMSE. Comparisons between the best ARIMA and ANN models showed that ANN outperformed ARIMA, with lower RMSE (5.5502) and MAE (4.1799). In conclusion, both ARIMA and ANN models effectively predicted API, however an ANN model offering slightly better performance based on the root mean square error.

Keywords*:* ARIMA, ANN, Air Pollution

Introduction

Pollution is a major global environmental issue, with air pollution being the leading cause of death. It consists of gases and particles, and its sources vary by location and time [1]. Air pollution is a growing concern for public health, especially for vulnerable populations such as children and the elderly [2]. In China, heavy industry and coal combustion contribute to air pollution, while in Australia, emissions from coal power stations exceed health safety standards [3,4].

The World Health Organization (WHO) reports that 92% of the global population faces poor air quality, and air pollution related deaths have increased [5]. Studies show that people in more polluted areas face higher mortality risks. Air pollution also negatively affects pregnancy and child development, leading to issues like low birth weight and neurodevelopmental disorders [6].

In Malaysia, air pollution is caused by rapid urbanization, industrialization and motor vehicle emissions [7]. Power plants contribute 85% of the pollution, with transboundary pollution from Indonesia's biomass burning also playing a significant role [8]. The Air Pollution Index (API) of Malaysia measures pollution using six pollutants, including sulphur dioxide (SO_2),

particulate matter (PM_{10}), ($PM_{2.5}$), ozone (O_3), nitrogen dioxide (NO_2) and carbon monoxide (CO) [9].

Klang Valley, a major focus of air pollution research in Malaysia, saw a significant increase in unhealthy days from 2022 to 2023 as shown in Figure 1 compared to other regions; Southern Regions, East Coast, Sabah, Federal Territory Labuan and Sarawak [9]. The region faces severe air quality issues due to industrial activity, traffic congestion and construction along with transboundary haze from forest fires in Indonesia [10,11]. This study aims to analyze the API trend in Klang Station, which has the worst air quality in Malaysia.

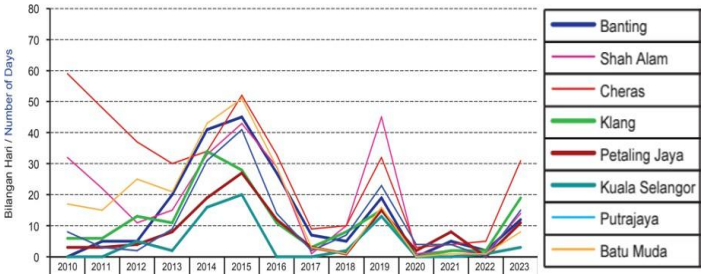


Figure 1: Number of Unhealthy API Days in Klang Valley in Year 2023
Sources: DOE, 2024.

Air pollution related deaths are rising annually. According to the Department of Statistics Malaysia (DOSM), pneumonia was the second leading cause of death in 2023, accounting for 13.3% of deaths, with a 2.2% increase from 2022. Pneumonia, an acute lung infection, inflames the tiny air sacs in the lungs, making it hard to breathe [12]. Children are particularly vulnerable, as they may struggle to fight off infections, leading to long term lung damage and respiratory issues [12].

In Malaysia, areas like Klang, with high API levels, are especially concerning. Klang's industrial activities, dense urbanization, and traffic emissions contribute to severe air pollution, linked to respiratory diseases like pneumonia. Klang's location is in the Klang Valley, surrounded by hills, trapped pollutants, and worsening air quality. Given these concerns, this study aims to develop a statistical model using Box Jenkins and ANN methods to identify the best-performing parameters and compare their effectiveness.

Methodology

The data used in this research were obtained from Malaysia's Department of Environment (DOE), covering the daily mean API and six major pollutants from 1 January 2019 to 31 December 2023. For the ARIMA model, API was used as the univariate input. In the ANN model, API is the output variable, with inputs including SO_2, PM_{10}, $PM_{2.5}$, O_3, NO_2 and CO, as listed in Table 1. Before undergoing analysis, the dataset underwent a data cleaning process and was then partitioned into 70% estimation and 30% evaluation. ARIMA model uses RStudio while ANN model will use JMP PRO17 to perform the result.

Table 1: List of Variables.

Variables	Unit
Sulphur Dioxide, SO_2	parts per million (ppm)
Particulate Matter, PM_{10}	microgram per cubic meter ($\mu g/m^3$)
Particulate Matter, $PM_{2.5}$	microgram per cubic meter ($\mu g/m^3$)
Ozone, O_3	parts per million (ppm)
Nitrogen Dioxide, NO_2	parts per million (ppm)
Carbon Monoxide, CO	parts per million (ppm)

Box Jenkins

The Box Jenkins methodology is a prominent approach for time series analysis, focusing on ARIMA models that combine Autoregressive (AR), Integrated (I) and Moving Average (MA) components. It includes three stages which are model identification, estimation and validation. The process begins with ensuring stationarity, as the Box Jenkins model assumes the data series is stationary. If this assumption is not met, steps must be taken to make the series stationary.

Stationarity is typically checked using the Unit Root Test, with the Augmented Dickey Fuller (ADF) Test being the most common due to its stable critical values, consistent performance across small and large sample sizes, and reliable power properties. The ADF Test operates under the null hypothesis (H_0) that the series has a unit root and is nonstationary, while the alternative hypothesis (H_1) assumes the series is stationary. A p-value less than 0.05 indicates stationarity.

If the data is not stationary, differencing is applied repeatedly until stationarity is achieved. ACF and PACF plots are then used to determine the model order (*p, d, q*). During estimation, models are evaluated for fit using metrics like AIC, BIC, and the Ljung Box test to ensure residuals are white noise. The selected model is then analyzed and refined to capture the underlying patterns and relationships in the data effectively [13].

Model Identification

The ACF and PACF plots help determine the Box-Jenkins model. An AR model is suitable when the ACF decays exponentially and the PACF has spikes, with the order (p) determined by significant PACF spikes. An MA model is preferred when the PACF decays and the ACF has spikes, with the order (q) based on notable ACF spikes. For irregular patterns in both plots, an ARMA model is used, with the order (p, q) corresponding to the spikes in the PACF and ACF, respectively [13]. The ARIMA model is expressed by equation 1 [13].

$$Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

where Y_t is Klang daily API, μ is constant term, $\phi_1, \phi_2, \ldots, \phi_p$ is an estimated parameters values. $\theta_1, \theta_2, \ldots, \theta_p$ is an estimated moving average parameters values, p is the autoregressive ordered, q is moving average ordered and ε_t refers to error term.

Model Estimation and Validations

Estimating models and conducting diagnostic tests are crucial for selecting the best forecasting model. Key objectives are ensuring the fitted values resemble actual values and using the fewest parameters for a good fit. Three common statistical measures for ARIMA models are the Ljung Box Test, Akaike's Information Criteria (AIC), and Bayesian Information Criteria (BIC). Firstly, Ljung Box Test checks if residuals are white noise. A p-value of the test less than 0.05 suggests model misspecification [13]. The null hypothesis of the Ljung Box Test is the errors are white noise.

Secondly, AIC evaluates model fit, with lower values indicating better fit. The model with the fewest parameters is preferred [13]. The equation 2 represents the AIC equation where k is the number of parameters in the model.

$$AIC = e^{\frac{2k}{T}} \left(\frac{\sum_{t=1}^{T} e_t^2}{T} \right) \tag{2}$$

Thirdly, BIC used to choose models by balancing model complexity and goodness of fit. The model with the lowest BIC is preferred, with a stronger penalty for more parameters [13]. The equation 3 represents the BIC where k denotes the number of parameters in the model, T refers to the number of observations.

$$BIC = T^{\frac{k}{T}} \left(\frac{\sum_{t=1}^{T} e_t^2}{T} \right) \tag{3}$$

Model Application

Once the test criteria are met and the model's fit is confirmed, it can generate forecast values. These forecasts are presented as single values. The model formulation and estimation process are repeated, refining it each time, until the best model is identified. The next step is to establish a system for monitoring forecasts. If new information shows the model is no longer accurate, it must be revised and updated.

Artificial Neural Network

An Artificial Neural Network (ANN) maps input variables (input layer) to output variables (output layer) through a learning process. It uses basic processors, called neurons, to identify relationships within data. Inspired by the biological structure of the human brain, ANNs are designed to learn, generalize, and make decisions [14].

ANNs are feed-forward neural network systems where neurons process inputs, recognize patterns, and perform nonlinear classification tasks [15]. Mathematically, an ANN can be represented as shown in equation 4 [16].

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \quad \text{and} \quad y_k = \phi(u_k + b_k) \tag{4}$$

Where bias, b_k, influences the net input of the activation function by either increasing or decreasing its value. The inputs are $x_1, x_2, \cdots, x_m$ and the weights of neuron k are $w_{k1}, w_{k2}, \cdots, w_{km}$. The linear combiner output resulting from the input signals is u_k. The activation function is denoted by $\phi(\bullet)$, and the output signal of the neuron y_k.

The ANN used was a multilayer network consisting of an input layer, an output layer, and a hidden layer in between. The hidden layer uses a transfer activation function to introduce non-linearity into the network. The activation of $a_j^{(l)}$ (the j^{th} neuron in the l^{th} layer) is related to the neurons in the (1-1)th layer by the equation 5. The equation 5, where $a_j^{(l-1)}$ is the k^{th} neuron in the (1-1)th layer, n_{l-1} is the total number of neuron in the (1-1)th layer, $w_{jk}^{(l)}$ is the weight for connection from the k^{th} neuron in the (1-1)th layer to the j^{th} neuron in the l^{th} layer $b_j^{(l)}$ is the bias of the j^{th} neuron in the l^{th}, f(*) is the activation function.

$$a_j^{(l)} = f\left(\sum_{k=1}^{n_{l-1}} w_{jk}^{(l)} a_k^{(l-1)} + b_j^{(l)} \right) \tag{5}$$

The input layer defines all the attribute values used as inputs for ANN model. In this study, the input values were SO_2, PM_{10}, $PM_{2.5}$, O_3, NO_2, and CO. The hidden layer consists of neurons that process the input data and pass the results to the output layer, with weights assigned to the input neurons to influence this process. The output layer contains neurons that represent the model's output, which in this study was the API.

The effectiveness of a model is evaluated using statistical measures that assess how well the observed outcomes are replicated. Two common indicators are the coefficient of determination, and the root mean square error [14]. Firstly, the coefficient of determination, R^2 is a widely used measure that indicates how well a model fits the data. An R^2 value closer to 1 suggests a better fit, as it means the model explains a greater proportion of the data's variability [17]. The equation 6 represent R^2 where $\hat{y}_i$ is the predicted i^{th} value, y_i is the i^{th} actual value, $\bar{y}$ is the mean of the actual value, m is the total number of observations.

$$R^2 = 1 - \left(\frac{\sum_{i-1}^{m} \left(y_i - y_i \right)^2}{\sum_{i-1}^{m} \left(y_i - \overline{y_i} \right)^2} \right) \tag{6}$$

Lastly, Root Mean Square Error, RMSE measures how well a model fits the data by quantifying the average magnitude of the prediction errors [18]. The equation 7 represents RMSE where n is the number of observations, y_i represents the actual value, $\hat{y}_i$ represents the predicted value.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{m}\left(y_i - y_i\right)^2} \tag{7}$$

Best Model Presentation

The performance of a model is evaluated using error measures, with lower values indicating better accuracy [13]. This study compared error measures, specifically Root Mean Square Error, (RMSE) and Mean Absolute Error (MAE), between the best ARIMA and ANN models. The RMSE evaluates performance by measuring the difference between predicted and observed values, with larger errors having a greater impact due to squaring [11]. A smaller RMSE indicates better model performance. The equation is shown in equation 7.

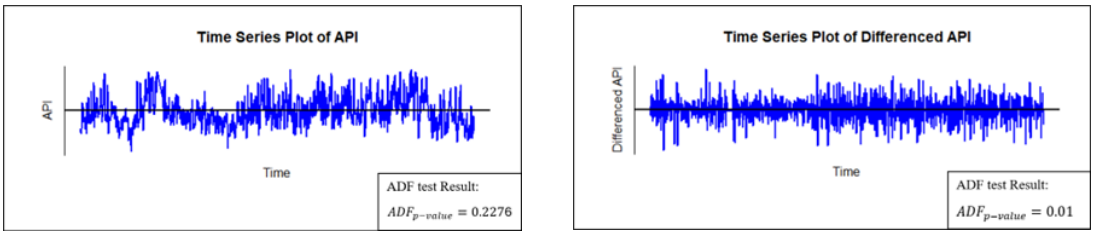
While, the MAE is less affected by large errors, making it stable, especially when actual values are near zero. A lower MAE indicates better model performance by showing closer average predictions to actual values [19]. The equation 8 represent MAE where n is the number of observations, y_i represents the actual value, $\hat{y}_i$ represents the predicted value.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{m}\left(y_i - y_i\right)^2} \tag{8}$$

Result and Analysis

Box Jenkins Methodology

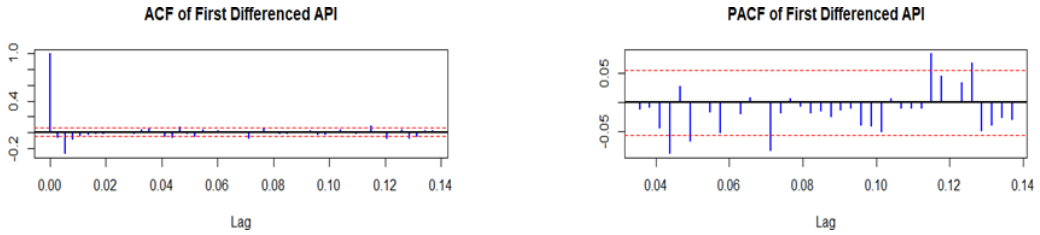
This study models daily API values using the ARIMA method in RStudio. Missing values and outliers were imputed. The Augmented Dickey Fuller (ADF) test is applied to ensure the time series data is stationary.



(a) before differencing (b) first differencing of API

Figure 2: API in Klang Valey.

Figure 2 illustrates the time series plots of API in Klang and the ADF test results. The initial plot in Figure 2(a) shows irregular fluctuations with the mean and variance of error values is not constant, and the ADF test showing p-value of 0.2276, indicating non-stationarity of the API series. After applying first differencing technique, the revised plot in Figure 2(b) displays consistent fluctuations with stable mean and variance. The ADF test confirms stationarity with a p-value of 0.01, supporting the rejection of the null hypothesis.



(a) ACF (b) PACF

Figure 3: API in Klang Valey.

Based on the ACF and PACF diagrams in Figure 3, significant spikes suggest potential ARIMA parameters. The ACF plot (Figure 3(a)) shows one significant spike at lag 2, suggesting an AR parameter of 1. Meanwhile, based on the PACF diagram (Figure 3(b)) gives five significant spikes (lags 4, 6, 11, 27 and 30), indicating an MA parameter of 5 were resulting in a potential ARIMA (5,1,1) model. Though, to identify the best model under ARIMA model [13], five other ARIMA models with parameters (p,d,q) = (4,1,1), (6,1,1), (5,1,2), (4,1,2) and (6,1,2) are tested. The Ljung Box test is used to assess each model for residual white noise, with the null hypothesis indicating no autocorrelation.

Table 2: Result of ARIMA Model.

Model	p-value	AIC	BIC
ARIMA (5,1,1)	0.2561	8292.351	8328.417
ARIMA (4,1,1)	0.2409	8290.440	8321.353
ARIMA (6,1,1)	0.2881	8294.070	8335.288
ARIMA (5,1,2)	0.2431	8294.387	8335.605
ARIMA (4,1,2)	0.2736	8292.210	8328.276
ARIMA (6,1,2)	0.2619	8296.337	8342.707

Table 2 shows the summary of the proposed ARIMA model result, which evaluates whether the errors in ARIMA models meet the white noise criteria. The p-values of all ARIMA models are greater than 0.05, leading to the rejection of the null hypothesis. Therefore, all six ARIMA models are well-specified and adequate, meeting the requirements for optimality. By evaluating the AIC and BIC, the ARIMA (4,1,1) model is identified as the best fit, with the lowest AIC and BIC values of 8290.44 and 8321.353, respectively. This indicates that the ARIMA (4,1,1) model strikes the best balance between model complexity and goodness of fit, outperforming the other candidate models. The ARIMA (4,1,1) model can be written as in equation 9.

$$\Delta_{y_t} = \phi_1 \Delta_{y_{t-1}} + \phi_2 \Delta_{y_{t-2}} + \phi_3 \Delta_{y_{t-3}} + \phi_4 \Delta_{y_{t-4}} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (9)$$

In the ARIMA (4,1,1) model, the observed value at time t is denoted as y_t, and the first difference of y_t is $\Delta y_t = y_t - y_{t-1}$. The model consists of four autoregressive (AR) parameters, ϕ_1, ϕ_2, ϕ_3, ϕ_4, which represent the influence of the previous four values on the current one. The differencing order, $d = 1$, is applied to make the data stationary by removing trends or patterns. The moving average (MA) parameter θ_1 accounts for the impact of the previous error term, which helps improve predictions by considering past prediction errors.

ANN Methodology

This study uses ANN to model API as the target variable, with SO_2, PM_{10}, $PM_{2.5}$, O_3, NO_2, and CO as inputs. Data screening and normalization were performed in JMP PRO 17 using imputation and the scale-and-offset method.

The ANN model comprises three layers: the input layer (receiving data), the hidden layer (assigning weights), and the output layer (target variable). To prevent overfitting, five networks with one to first hidden nodes were constructed. Model performance was evaluated using the coefficient of determination (R^2) and root mean square error (RMSE). The best-fit model had the highest R^2 and lowest RMSE, subject to model parsimony. Table 3 presents the ANN model results.

Table 3: R^2 and RMSE Result.

Model	Hidden Nodes	AIC	BIC
1	1	0.6102	5.8829
2	2	0.6165	5.8347
3	3	0.6184	5.8209
4	4	0.6306	5.7269
5	5	0.6354	5.6894

Model 5, with five hidden nodes, delivers the best performance, achieving the highest R^2 value of 0.6487 and the lowest RMSE of 5.7284 as shown in Table 3. This indicates that the model effectively captures relationships in the data, explains variability, and provides accurate predictions. The model also shows low error rates on both training and testing datasets, indicating a balance between learning from the data and generalizing to unseen data.

This suggests that Model 5 avoids overfitting, where a model performs well on training data but poorly on new data, and underfitting, where it fails to learn the patterns in the training data. The five hidden nodes represent processing units in the hidden layer that work together to identify patterns and relationships between the input variables and the target, API. These nodes, with their respective weights and biases, help the model generalize well, making Model 5 suitable for forecasting API values in Klang. Figure 4 illustrates the network structure of Model 5.

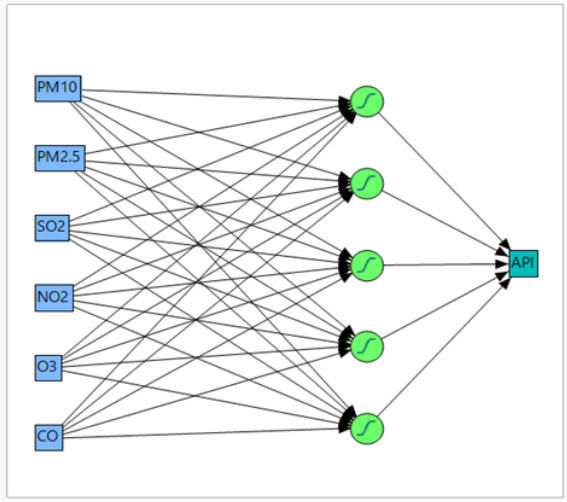


Figure 4: Model 5 Network.

The Model Performance

Table 4 shows that ANN Model 5, with five hidden nodes, outperforms ARIMA (4,1,1) based on error measures. ANN achieves an RMSE of 5.5502 and an MAE of 4.1799, compared to ARIMA's RMSE of 14.35176 and MAE of 11.28785. This demonstrates that ANN provides more accurate predictions and is better at capturing patterns in the data.

Table 4: R^2 and RMSE Result.

Model	RMSE	MAE
ARIMA (4,1,1)	14.35176	11.28785
ANN Model 5	5.5502	4.1799

Conclusion and Recommendation

The study applied both the Box-Jenkins and ANN methods to forecast API values. The Box-Jenkins approach identified ARIMA (4,1,1) as the best model based on its ability to balance accuracy and model complexity, demonstrated by the lowest AIC and BIC values. On the other hand, the ANN method determined that Model 5, with five hidden nodes, provided the best performance. This model achieved the highest R^2 and the lowest RMSE, indicating its effectiveness in capturing data patterns and maintaining low error rates. Overall, the ANN method showed superior performance compared to the Box-Jenkins method due to its lower error measurements and better adaptability to the data.

Future research could benefit from exploring additional modeling techniques beyond Box-Jenkins and ANN, such as linear regression, logistic regression, exponential smoothing, or Holt's method. Incorporating a wider range of methods could uncover unique patterns in the data and provide more insights. Additionally, using larger and more recent datasets could enhance the model's reliability and its ability to generalize across various environmental

conditions. This would also help researchers evaluate the model's performance over different time periods, improving its effectiveness for forecasting air quality.

Lastly, future studies should focus on real-time forecasting of API values, which was not part of the current research. Utilizing real-time air pollution data could lead to more accurate predictions and allow for better monitoring of daily air quality. Providing timely warnings about poor air conditions could help mitigate the health risks associated with air pollution and improve public health outcomes.

Acknowledgments

The researcher would like to sincerely thank everyone who helped make this study a success. Special thanks go to the management of Universiti Teknologi MARA Seremban Campus for their support and help in making this research possible.

References

[1] Susanto, A. D. (2020). Air pollution and human health. *Medical Journal of Indone- sia*, 29, 8–10. https://doi.org/https://doi.org/10.13181/mji.com.20457.

[2] Karliansyah, M. (2024). View of air pollution impacts on human health and policies to reduce air pollution *Medical Journal of Indonesia* https://mji.ui.ac.id/journal/ in- dex.php/mji/article/view/4579/1723Chen, J., & Mu, X. (2021). Review on environmental treatment of heavy pollution industry. https: // doi.org / https: //doi.org /10.1051 /e3sconf / 202132901045

[3] Roden, J. (2020). Air pollution issues in australia. *Medical Research Archives* 8. https: //doi.org/https://doi.org/10.18103/mra.v8i10.2262

[4] WHO. (2016) WHO releases country estimates on air pollution exposure and health impact

[5] WHO. (2023). Air Pollution

[6] Chang, F., & Ashfold, M. J. (2020). Public perceptions of air pollution and its health impacts in greater kuala lumpur. *IOP Conference Series: Earth and Environmental Science* 489. https://doi.org/https://doi.org/10.1088/1755-1315/489/1/012027

[7] Sentian, J., Herman, F., Yih, C. Y., & Hian Wui, J. C. (2019). Long-term air pol- lution trend analysis in malaysia. *International Journal of Environmental Impacts: Management, Mitigation and Recovery, 2*. DOI=https://doi.org/10.2495/ei-v2-n4- 309-324

[8] DOE. (2024). Environmental quality report 2023. Department of Environment Malaysia.

[9] Sahak, N., Asmat, A., & Yahaya, N. Z. (2022). Spatio-temporal air pollutant characterization for urban areas. *Journal of Geoscience and Environment Protection, 10(01)*, 218–237. https://doi.org/https://doi.org/10.4236/gep.2022.101015

[10] Albashir Abdulali, B. A., & Masseran, N. (2021). Artificial Neural Network (ANN) and ARIMA Models for Better Forecast of the Air Pollution Data in Malaysia.

[11] WHO. (2021). Pneumonia. https://www.who.int/news-room/fact-sheets/detail/pneumonia

[12] Lazim, M. A. (2011). *Introductory business forecasting : A practical approach* (3rd ed). UiTM Press.

[13] Mousavian, A., Mahmodabady, H. Z., & Jadval Ghadam, A. G. (2015). Prediction of the pollutants concentration using artificial neu- ral network (ann). *Environment Conservation Journal, 16(SE)*, 171–180. https://doi.org/https://doi.org/10.36953/ecj.2015. se1619

[14] Sobri, N. M., Yaacob, W. F. W., Ismail, N. A., Malik, M. A. A., Rahman, R. A., Baser, N. A., & Sukhairi, S. A. M. (2021). Predicting particulate matter (pm2.5) in malaysia using multiple linear regression and artificial neural network. *Jour- nal of Physics: Conference Series.* https://doi.org/https://doi.org/10.1088/1742- 6596/2084/1/ 012010

[15] Ku Yusof, K. M. K., Azid, A., Abdullah Sani, M. S., Samsudin, M. S., Muham- mad Amin, S. N. S., Abd Rani, N. L., & Jamalani, M. A. (2019). The evalua- tion on artificial neural networks (ann) and multiple linear regressions (mlr) mod- els over particulate matter (pm10) variability during haze and non-haze episodes: A decade case study. *Malaysian Journal of Fundamental and Applied Sciences.* https://doi.org/https: //doi.org/ 10.11113/ mjfas.v15n2.1004

[16] Zulkepli, N. E. S. (2021). A study on air pollution index in sabah and sarawak using principal component analysis and artificial neural network. http : / / ir. uitm . edu . my/id/ eprint/46511/1/46511.pdf

[17] Maltare, N. N., & Vahora, S. (2023). Air quality index prediction using machine learning for ahmedabad city. 7, 100093–100093. https://doi.org/https://doi.org/10.1016/j.dche.2023.10 0093

[18] Liu, T., & You, S. (2022). Analysis and forecast of beijing's air quality index based on arima model and neural network model. Atmosphere, 13. https : // doi . org / https: //doi.org/ 10.3390/atmos13040512

PADDY PRODUCTION ANALYSIS IN NON-GRANARY AREAS

Nora Binti Mohd Basir[1*], Nur Aliah Najwa Binti Ahmad Mahyudin[1]
[1]*College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan*
[*]noram661@uitm.edu.com

Abstract

Currently, it has been shown that Malaysia has a low production of paddy. That is one of the reasons why Malaysia has been importing rice from other countries. To help Malaysia overcome the rice output shortage, this study has decided to determine whether planted area and paddy yield affect paddy production in a long-run scenario. This study analyses the paddy production in non-granary areas in Johor, Melaka, Negeri Sembilan, Pahang, Pulau Pinang, and Terengganu from 2011 until 2020, during the main season in Malaysia. In order to determine the presence of the long-run relationship, this study employed the Pedroni cointegration test. It was found that most of the statistical tests from the Pedroni cointegration test are significant at a 5% significance level. It can be concluded that there is a long-run relationship between planted area, paddy yield, and paddy production. Thus, it is recommended that future researchers investigate these effects in detail to help maximise paddy production, especially in non-granary areas.

Keywords*:* Cointegration test, Paddy production, Panel data, Stationary test

Introduction

The world's three most important food crops are rice, wheat, and maise. When taken as a whole, rice makes up almost half of the calories people consume worldwide [9]. Besides, rice has also been one of the staple foods in many countries around the world. Countries located at the Association of Southeast Asian Nations (ASEAN), also known as the Southeast Asian regions, have been categorised as the countries that consume rice the most. In 2023, a study found that Malaysia only consumes 106 kilograms of rice per person yearly, which is lower than some Southeast Asian countries [14]. Table 1 shows Malaysia's total rice consumption from 2019 to 2024 [10].

Table 1: Total Consumption of Rice in Malaysia

Year	Rice Consumption (million metric tons)
2019	2.75
2020	2.85
2021	2.90
2022	2.90
2023**	2.91
2024**	2.91

** indicates the estimation of rice consumption in the future.

In 2021, Malaysia only produced roughly 1.68 million metric tons of rice, 180 thousand metric tons more than the production in 2020 [11]. It shows that rice production does not align with rice consumption as the current production is still unable to meet the demands of the Malaysian market. Naturally, as the demand for rice increases, production should also rise to correspond to the demand. While Malaysia produces 67% of its rice, the remaining is imported from Pakistan, Vietnam, and Thailand [12]. Thus, the Malaysian government is preparing to increase local rice production by 2025 [11].

Essentially, Malaysia needs to produce more rice. To maximise production, Malaysia should focus on land and climate change. Malaysia has two seasons: main season and off-season, during which the activities of paddy cultivation, from the preparation of land to harvesting, start [6]. Table 2 shows that the non-granary areas in Johor, Melaka, Perak, and Terengganu have a decreasing productivity status. Meanwhile, Kedah, Negeri Sembilan, and Perlis have optimum productivity, while productivity in Kelantan and Pahang have an increasing pattern.

Table 2: Summary of Productivity Status

Non-Granary States	Productivity Status	Non-Granary States	Productivity Status
Johor	Decreasing	Pahang	Increasing
Kedah	Optimum	Perak	Decreasing
Kelantan	Increasing	Perlis	Optimum
Melaka	Decreasing	Terengganu	Decreasing
Negeri Sembilan	Optimum		

Many aspects should be considered when discussing the causes of paddy production. A study found that rice production can be affected by human capital, labour, wages, wetlands, urban population, and rice prices [3]. However, the technology has barely any effect on the production. Exploring both short- and long-run factors affecting paddy production is important. A study on rice production in Nepal revealed that in both short- and long-term, rice production is positively influenced by cultivated area, fertiliser consumption, and agricultural credit [5].

The average yield of cleaned paddy is the mean crop after drying and cleaning and when the moisture content is precisely 14% [6]. There is no significant difference in the yield during the dry season, regardless of the types of water management [13]. On the contrary, the yield differs with alternate wetting and drying systems having higher paddy yields than continuous flooding [13]. Next, [6] stated that the paddy plantation is divided into granary and outside granary areas. The granary areas refer to the dominant irrigation schemes, which are about 4,000 hectares minimum, while the outside granary areas are those with minor or outside irrigation systems.

A study by [13] used both water management systems, continuous flooding (CF) and alternate wetting and drying (AWD), to manage water in maintaining rice production. During the wet season, CF needed more irrigation than AWD, as CF irrigates 1,835 mm of water while AWD irrigates 1,522 mm [13].

Nevertheless, the current paddy production is insufficient to cover Malaysia's population. Since the granary areas are likely to have optimum and increasing rice productivity, this study focuses on the non-granary areas during the main season to help increase productivity. Hence,

the study will determine whether a long relationship exists between planted area, paddy yield, and paddy production.

Methodology

Source of Data

This study used data from the Paddy Production Survey Report 2020/2021 by the Department of Agriculture, Peninsular Malaysia. The data provided is categorised as panel data, a combination of cross-sectional and time series data [15]. The annual data utilises information from six states in Malaysia, specifically non-granary states, from 2011 until 2020, with 60 observations and two independent variables. The six non-granary states are Johor, Melaka, Negeri Sembilan, Pahang, Pulau Pinang, and Terengganu, and the independent variables that are being investigated are the planted area and paddy yield. Table 3 shows the variable of interest in this study.

Table 3: List of Variables

Variable	Description	Measurement
AREA	Planted area in hectare	Ratio
YIELD	Average yield of cleaned paddy in kilogram/hectare	Ratio
PRODUCTION	Wetland paddy production in metric ton	Ratio

Stationary Test

Since panel data combines time-series and cross-sectional data, the study needs to test for model stationary using a unit root test. The study decided to use the Hadri Lagrange Multiplier test proposed by Hadri in 2000. Hadri has a null hypothesis of stationary, while in contrast, the alternative hypothesis of at least one series has a unit root [7]. The Hadri test's components are represented by Equation 1 [1].

$$y_{it} = z_{it}'\gamma + r_{it} + \varepsilon_{it} \tag{1}$$

Where z_{it} is the deterministic component, while r_{it} is a random walk,

$$r_{it} = r_{it} + u_{it} \tag{2}$$

And ε_{it} is known as a stationary process, while $u_{it} \sim iid(0, \sigma_u^2)$. The partial sum of the residuals, S_{it} is written as in Equation 3

$$S_{it} = \sum_{j=1}^{t} \hat{e}_{ij} \tag{3}$$

While the LM statistic represented in Equation 4

$$LM = \frac{\frac{1}{n}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T} S_{it}^2}{\hat{\sigma}_{\varepsilon}^2} \tag{4}$$

Hadri (2000) believes that the standardised statistics are as in Equation 5 and Equation 6 [7]:

$$Z_{\mu} = \frac{\sqrt{N}(LM_{\mu} - \xi_{\mu}}{\zeta_{\mu}} \Rightarrow N(0,1) \tag{5}$$

and

$$Z_{\tau} = \frac{\sqrt{N}(LM_{\tau} - \xi_{\tau}}{\zeta_{\tau}} \Rightarrow N(0,1) \tag{6}$$

As $T \rightarrow \infty$ followed by $N \rightarrow \infty$.

Cointegration Test

To verify the stationary variables, a unit root test is run [2]. It is crucial to determine whether the stationary first-difference variables, $I(1)$, are cointegrated. The cointegration test also checks whether a long-run relationship among the variables exists or not [4]. The study uses Pedroni's (1999) residual-based cointegration test to test the cointegration. It is constructed based on Engle-Granger (1987), as the test statistic is from the residuals of the panel static regression [2].

The test statistics and p-values by Pedroni (1999) are a panel and weighted panel of v-statistic, ρ-statistics, PP-statistic, and ADF-statistic. Besides, there are also group mean approaches of ρ-statistics, PP-statistic, and ADF-statistic [8]. The panel and weighted panel statistics are from pooled data along the within-dimension, while the group statistics are from pooled data along the between-dimension [2]. Pedroni had designed the following test statistics [2]:

1. Panel v-statistic:

$$Z_{\hat{v}_{NT}} = \frac{1}{\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{L}_{11i}^{-2}\ \hat{e}_{it-1}^{2}\right)} \tag{7}$$

2. Panel ρ-statistic:

$$Z_{\hat{\rho}_{NT}-1} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{L}_{11i}^{-2}\left(\hat{e}_{it-1}\Delta\hat{e}_{it} - \hat{\lambda}_i\right.}{\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{L}_{11i}^{-2}\ \hat{e}_{it-1}^{2}\right)} \quad (8)$$

3. Panel t-statistic (non-parametric):

$$Z_{t_{NT}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{L}_{11i}^{-2}\left(\hat{e}_{it-1}\Delta\hat{e}_{it} - \hat{\lambda}_i\right)}{\sqrt{\tilde{\sigma}_{NT}^{2}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{L}_{11i}^{-2}\ \hat{e}_{it-1}^{2}\right)}} \quad (9)$$

4. Panel t-statistic (parametric):

$$Z_{t_{NT}}^{*} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{L}_{11i}^{-2}\ \hat{e}_{it-1}^{*}\Delta\hat{e}_{it}^{*}}{\sqrt{\tilde{s}_{NT}^{*2}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{L}_{11i}^{-2}\ \hat{e}_{it-1}^{*2}\right)}} \quad (10)$$

5. Group ρ-statistic:

$$\tilde{Z}_{\hat{\rho}_{NT}-1} = \sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\left(\hat{e}_{it-1}\Delta\hat{e}_{it} - \hat{\lambda}_i\right)}{\left(\sum_{t=1}^{T}\hat{e}_{it-1}^{2}\right)} \quad (11)$$

6. Group t-statistic (non-parametric):

$$\tilde{Z}_{t_{NT}} = \sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\left(\hat{e}_{it-1}\Delta\hat{e}_{it} - \hat{\lambda}_i\right)}{\sqrt{\hat{\sigma}_i^{2}\left(\sum_{t=1}^{T}\hat{e}_{it-1}^{2}\right)}} \quad (12)$$

7. Group t-statistic (parametric):

$$Z_{t_{NT}}^{*} = \sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\hat{e}_{it-1}^{*}\Delta\hat{e}_{it}^{*}}{\sqrt{\sum_{t=2}^{T}\tilde{s}_i^{*2}\hat{e}_{it-1}^{*2}}} \quad (13)$$

As mentioned before, the null hypothesis of the Pedroni residual cointegration test is that there is no cointegration relationship between the non-stationary variables. So, to estimate a model in a long-run relationship, the study must reject the null hypothesis. To reject the null hypothesis, the study must ensure that the majority of the test statistics and their p-value are significant. Only then can it be proven that there is a long-run relationship between planted area, paddy yield and paddy production.

Results and Discussion

Stationary Test

Since panel data combines cross-sectional and time series data, stationary assumptions must be checked. This study uses the Hadri Lagrange Multiplier (LM) Stationary test, as stated in Table 4.

Table 4: Hadri LM test

Variables	Level	First Difference
AREA	1.443e-03 *	0.5037
YIELD	1.203e-06 *	0.7009
PRODUCTION	1.035e-04 *	0.6687

* significant at 0.05.

Table 4 displays that at level, all *p*-values are significant at 5%. The study can conclude that area, yield, and production are not stationary at this level. Hence, the variables must be tested again at the first difference level. Next, it is proven that area, yield, and production are stationary at first difference as the *p*-values are insignificant. It can be concluded that all variables are stationary at first difference and integrated at first order, $I(1)$. Thus, the study can check the cointegration between the variables to determine the existing long-run relationship between them.

Cointegration Test

A cointegration test is performed since the study wants to determine the long-run relationship between planted area, paddy yield, and paddy production. If cointegration exists, it can be proven that there is a long-run relationship.

Table 5: Pedroni Residual Cointegration test

	Test	Statistic	Probility
Alternative hypothesis: common AR coefs. (within-dimension)	Panel v-statistic	-1.2141	0.8877
	Panel rho-statistic	1.1603	0.8770
	Panel PP-statistic	-1.7310	0.0417 *
	Panel ADF-statistic	-1.7291	0.0419 *
	Weighted Panel v-statistic	-1.5085	0.9343
	Weighted Panel rho-statistic	0.7018	0.7586
	Weighted Panel PP-statistic	-6.9495	0.0000 *
	Weighted Panel ADF-statistic	-5.4273	0.0000 *
Alternative hypothesis: common AR coefs. (between-dimension)	Group rho-statistic	1.6467	0.9502
	Group PP-statistic	-7.6406	0.0000 *
	Group ADF-statistic	-4.8891	0.0000 *

* significant at 0.05.

Six out of eleven *p*-values from Table 5 are less than 0.05. Hence, it can be concluded that the variables are cointegrated. Thus, the study can conclude that there is a long-run relationship between planted area, paddy yield, and paddy production.

Conclusion and Recommendation

Rice is the staple food in Malaysia, and it is common for the demand for rice to exceed that of other foods. When the demand for rice increases, paddy production should ideally increase proportionately to meet this demand. However, since Malaysia's market needs are not being fulfilled, this study explores the factors that could enhance local paddy production. The dataset of paddy production, planted area, and paddy yield during the main season from 2011 until 2020 for a few non-granary states in Malaysia was analysed. Hadri LM test is applied to test the stationary, and it is found that planted area, paddy yield and paddy production are not stationary at level but stationary at first difference. Hence, the Pedroni residual cointegration test is run. From the cointegration test, the finding reveals that all the variables are cointegrated with each other. It is concluded that there is a long-run relationship between planted area, paddy yield, and paddy production.

Hence, the study believes that there is a long-run effect between planted area and paddy yield with paddy production. Therefore, future researchers should investigate in detail these effects and how the planted area and paddy yield affect the paddy production output in the long run. Besides, it is recommended for future studies to take into account the data around the whole year, which includes both main and off seasons, since this research only investigates during the main season.

Acknowledgement

We sincerely thank each member for their unwavering support and commitment, which ensured that this article was published successfully.

References

[1] Baltagi, B. H., & Kao, C. (2000). Non-stationary panels, cointegration in panels and dynamic panels: A survey. *Advances in Econometrics*

[2] Barbieri, L. (2008). Panel Cointegration Tests: A Survey. *Rivista Internazionale Di Scienze Sociali, 116*(1), 3–36. https://www.jstor.org/stable/41625199

[3] Bashir, A., & Yuliana, S. (2019). Identifying factors influencing rice production and consumption in Indonesia. *Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi Dan Pembangunan, 19*(2) https://doi.org/10.23917/jep.v19i2.5939

[4] Bouchoucha, Yahyaoui, I., & Bouchoucha, N. (2019). The long-run relationship between ODA, growth and governance: An application of FMOLS and DOLS Approaches

[5] Chandio, A. A., Jiang, Y., Ahmad, F., Adhikari, S., & Ain, Q. U. (2021). Assessing the impacts of climatic and technological factors on rice production: Empirical evidence from Nepal. *Technology in Society, 66* 101607. https://doi.org/10.1016/j.techsoc.2021.101607

[6] Department of Agriculture Peninsular Malaysia (2021). Paddy Production Survey Report MALAYSIA, Main Season 2020/2021.

[7] Giulietti, M., Otero, J., & Smith, J. (2009). Testing for stationarity in heterogeneous panel data in the presence of cross-section dependence. *Journal of Statistical Computation and Simulation, 79*(2), 195–203. https://doi.org/10.1080/00949650701719136

[8] Örsal, D. D. K. (2007). Comparison of panel cointegration tests. Www.econstor.eu. http://hdl.handle.net/10419/25201

[9] Pariona, A. (2019). What are the World's Most Important Staple Foods? WorldAtlas. https://www.worldatlas.com/articles/most-important-staple-foods-in-the-world.html

[10] Statista. (2023a). Annual Rice Consumption in Malaysia 2019-2024. https://worldpopulationreview.com/country-rankings/rice-consumption-by-country

[11] Statista. (2023b). Production of Rice in Malaysia 2013-2021. https://www.statista.com/statistics/794700/rice-production-volume-malaysia/

[12] Syahmi Makhtar, Irwan Shah Zainal Abidin, & Rabiul Islam. (2022). REVIEWING FOOD SECURITY ON PADDY PRODUCTION: A CONCEPTUAL PAPER. *International Journal of Industrial Management, 16(*1), 51–58. https://doi.org/10.15282/ijim.16.1.2022.9012

[13] Win, E. P., Win, K. K., Bellingrath-Kimura, S. D., & Oo, A. Z. (2020). Greenhouse gas emissions, grain yield and water productivity: a paddy rice field case study based in Myanmar. *Greenhouse Gases: Science and Technology, 10*(5), 884–897. https://doi.org/10.1002/ghg.2011

[14] World Population Review. (2024). Rice Consumption by Country 2023. Worldpopulationreview.com. https://worldpopulationreview.com/country-rankings/rice-consumption-by-country

[15] Yohandoko, S., & Supriyanto, S. (2023). Panel Data Analysis on Rice (Paddy) Production in Indonesia 2018-2021. *International Journal of Mathematics, Statistics, and Computing, 1*(3), 44–53. https://doi.org/10.46336/ijmsc.v1i3.27

HOME ENERGY MANAGEMENT FOR HOME APPLIANCES USING KNAPSACK MODEL

Nor Aishah Md Noh, Fatin Nur Liyana Mohd Ridzuan
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban,
Negeri Sembilan, Malaysia.
*nor_aishah@uitm.edu.my

Abstract

In recent months, increasing electricity bills have led to the need for consumers to manage their energy consumption efficiently. This study used the knapsack model for home energy management for home appliances to solve this problem. The aims are to adapt an optimization approach for daily appliance usage using the knapsack model for home energy management of home appliances and to compare the cost of electricity bills per day before and after using the knapsack model. The model was developed using IBM CPLEX software where the knapsack model will prioritize essential appliances for continuous use while nonessential devices are scheduled based on the cost budget limit. The results indicate that the knapsack model reduced daily electricity costs by RM0.22–RM0.66. This corresponds to monthly savings of RM6.60–RM19.80, representing approximately 4–12% compared to the actual daily costs. In conclusion, the model not only ensures optimal energy consumption but also provides consumers with a sustainable and cost-effective solution for managing household electricity usage.

Keywords*:* Energy consumption, appliances, knapsack model, IBM CPLEX

Introduction

The knapsack model is a combined optimization method where the goal is to maximize the total value of the items put into the bag without exceeding the weight of the bag [1]. Researchers often use the knapsack model in their studies, especially in solving optimization problems most efficiently. It shows that in home energy management (HEM), the knapsack model can optimize energy consumption (EC) to effectively meet household demand. This is one of the ways of HEM to maximize the EC in the home by using the knapsack model. HEM is the practice of managing EC in the home effectively and efficiently to reduce the cost of EC [2]. The cost of EC needs to be controlled because, in recent months, an increase in the price of electricity bills has been an issue that is getting more attention in the community. Consumers need to understand the factors that lead to this increase, as it helps manage their budget successfully while minimizing the impact on their quality of life.

The management of household EC has become a crucial issue in the modern period, due to the growth of home appliances for daily functions [3]. Efficient EC not only lowers utility expenses but also helps protect the environment by lowering carbon emissions. HEM has some difficulties in terms of assigning weights, since various appliances have different energy consumption levels appliances have different weights. That means managing appliances with varying weights complicates the optimization challenge. Determining the right weights needs

detailed information about each appliance's EC, consumption patterns, and significance to the home, which can be difficult to gather perfectly. The knapsack model can help solve EC problems which are often used for optimization and offers a systematic way to divide limited resources for maximum benefit. In HEM, these models help households prioritize EC for different appliances based on immediate needs, consumer preferences, and energy efficiency ratings. However, this knapsack also has its weaknesses. The data used must be in a large amount to ensure that the result made is more accurate [4].

The objective of this study is to adapt an optimization usage of appliances per day using the knapsack model for home energy management of home appliances. Next, to compare the cost of electricity bills per day before and after using the knapsack model. This study focuses on the knapsack model to address HEM for home appliances in urban areas of Malaysia. This study focuses on urban Malaysia, where EC patterns differ from rural areas [5]. Besides, instead of analyzing EC weekly, this study will do it daily to produce a more accurate and detailed representation of consumption patterns which include increases and decreases in energy demands and wants. The daily analysis enables easier optimization and ensures that energy efficiency closely corresponds with the lifestyle of urban residents.

The significance of this study comes from its potential to address critical issues related to energy efficiency, cost savings and sustainability in households. EC accounts for a large portion of household expenses, so it needs to be managed properly so that the price of the electricity bills can be reduced. It is very important to implement an effective HEM strategy to minimize energy waste and maximize energy efficiency. The use of the knapsack model helps consumers reduce their electricity bills by optimizing the scheduling and operation of appliances based on factors such as EC rates and tariff structures. Carbon emissions and environmental degradation are the effects of household EC. Households can reduce their carbon emissions and contribute to environmental sustainability by optimizing EC with the knapsack model.

Literature Review

The knapsack model is a well-known optimization method for dealing with multiple issues. It is widely used in situations where a group of objects with related weights and values must be chosen to maximize the overall value while keeping to a total weight limit [6]. The main purpose of the knapsack model is to determine the maximum amount of profit that can be brought with the maximum capacity of the knapsack. Each appliance has its own weight and profit [7]. This study used the knapsack model because it offers several advantages over other optimization algorithms. The knapsack model helps to achieve optimal consumption of limited resources by selecting appliances that maximize the overall value within the cost budget. Next, the basic idea of the knapsack model is easy to understand. Beginners and experienced users can use it because of its simplicity [8]. By integrating the knapsack model into home energy management (HEM), energy consumption (EC) can be optimized without exceeding the cost budget.

Besides the knapsack model, other optimization methods such as the greedy algorithm and brute force algorithm have also been used in problem solving. The greedy algorithm makes decisions step by step by choosing what looks most profitable at that moment. This method is simple and fast. However, it only provides locally optimal solutions and does not always guarantee the best overall outcome, which limits its suitability for energy

management problems [4-9]. The brute force algorithm, on the other hand, tries all possible solutions systematically to find the best result. Although it ensures accuracy, it is highly resource-demanding and inefficient for large datasets. It is because it relies heavily on computing power [4]. The knapsack model offers a more balanced approach compared to these methods. It requires less computation than brute force and produces better-quality solutions than the greedy algorithm. These strengths make the knapsack model both practical and effective for scheduling household appliances within a limited cost budget.

HEM is the process of planning, managing, and controlling the EC in a house efficiently and effectively. It involves the use of methods that aim to reduce energy waste, increase awareness of EC, and reduce the negative impact on the environment. The purpose of HEM is to schedule household appliances to reduce electricity bills [10]. Electricity bills have increased because the increased mobility contributes to higher EC, with people using more electrical devices in different places [11]. Consumers can use renewable energy at home to reduce electricity bills. However, it is difficult to install it, especially for the B40 group [12], but it is not an obstacle for consumers to reduce the cost of their electricity bills. Consumers can use various other methods to reduce their electricity bills such as unplugging devices and turning off the standby mode. Consumers can also set EC prioritize and limits for each appliance based on factors such as immediate needs, consumer preferences, and energy costs.

Methodology

Data Collection Method

Primary data was obtained through a questionnaire conducted using Google Forms. The questionnaire was used as a survey instrument to obtain the socioeconomic profile of households and the type of electrical appliance consumption. This study will use a convenience sampling technique where the selected respondents are from 18 years old to 50 years old and live in urban areas. This sampling technique can minimize selection bias and be representative of the population [13]. The sample of 30 respondents used for this study is all Malaysians in all states who live in urban areas. This study collects 30 respondents as a sample due to cost limitations and the time required to collect data from a larger population. The sample was considered sufficient to obtain an initial view of household energy usage within the scope and resources available for this study. Table 1 shows the data about the percentage of respondents who have those appliances out of 30 respondents and the average duration of usage per day.

Table 1: Summary of the respondents' appliances used with average usage time from the survey

Appliances Parameter (i)	**Description of Appliances**	**Appliances' User (% out of 30 respondents)**	**Average Usage Time (hours per day)**
1	Refrigerator	100	24
2	Fan	100	24
3	Lamp	100	12
4	Charger Device	100	3

Continue to the next page

Continue from the previous page			
5	Washing Machine	100	2
6	Iron	100	1
7	Rice Cooker	86.67	1
8	Television	83.33	5
9	Freezer	66.67	24
10	Air Fryer	56.67	1
11	Vacuum	53.33	1
12	Water Purifier	50	24
13	Wi-Fi	50	24
14	Pressure Cooker	50	1
15	Air Conditioning	43.33	8
16	Microwave	23.33	1

For secondary data, the data were obtained from previous studies and used the power rating information. The data used is obtained from the journal [8-14] and Tenaga Nasional Berhad (TNB). Table 2 shows data about the power rating for each appliance, the duration of usage for each appliance, and a list of all appliances to be used. The duration of usage is taken from the average time the respondents use the appliances daily. The cost of appliances is calculated using (1) and the cost for each time slot is RM0.198.

$$C_i = \sum_{i=1}^{n} P_i \, x_i \, C_S \tag{1}$$

where,

C_i = Cost of appliances i (RM/hour)
i = Appliance
n = The total number of appliances
P_i = The power rating of the appliances i (kilowatts)
x_i = The duration of usage for the appliances i per day (hours)
C_S = The cost for each time slot (RM/kWh)

Table 2: Initial parameter for knapsack appliances

Appliances Parameter (i)	**Description of Appliances**	**Power Rating (kilowatts)**	**Duration of Usage (hours per day)**	**Cost of Appliances, C_i (RM per hour)**
1	Refrigerator	0.400	24	0.0792
2	Fan	0.054	24	0.0107
3	Lamp	0.030	12	0.0059
Continue to the next page				

Continue from the previous page				
4	Charger Device	0.005	3	0.0010
5	Washing Machine	0.425	2	0.0842
6	Iron	1.000	1	0.1980
7	Rice Cooker	0.300	1	0.0594
8	Television	0.100	5	0.0198
9	Freezer	0.140	24	0.0277
10	Air Fryer	1.400	1	0.2772
11	Vacuum	0.600	1	0.1188
12	Water Purifier	0.300	24	0.0594
13	Wi-Fi	0.006	24	0.0012
14	Pressure Cooker	1.000	1	0.1980
15	Air Conditioning	0.750	8	0.1485
16	Microwave	0.700	1	0.1386

Implementation of the Knapsack Model

The knapsack problem is an optimization model used to determine the best appliance combination based on weight and value. The objective is to maximize the total value while ensuring the total weight does not exceed the capacity limit. The objective can be formulated as (2). By applying the knapsack model, this study optimizes home energy management by prioritizing appliances that provide the most benefit within the given energy constraints.

$$\text{Maximize } E = \sum_{i=1}^{n} v_i \, x_i \tag{2}$$

$$\text{Subject to } = \sum_{i=1}^{n} w_i \, x_i \leq W$$

where,

E = Total energy consumption per day
i = Appliance
n = Total number of appliances
v_i = Energy consumption of appliances i for each hour
x_i = Duration of usage for the appliances i per day
w_i = Cost of appliances i for each hour
W = Cost budget per day

Everyone has different home appliances based on their needs, lifestyle, and financial capabilities. These differences cause different energy consumption (EC) for each home. Therefore, this study took three respondents as a case to represent the variation in the use of home appliances. These respondents were selected based on specific criteria, such as their energy usage patterns and appliance ownership to provide useful information about the impact

of the knapsack model on home energy management (HEM). The goal of these cases is to analyze the variety of possibilities regarding the duration of usage of the appliances per day, as well as the uses for those cases that are subject to constraints. Table 3 shows the list of available appliances based on respondents' feedback. This sample selection allows for a deeper analysis of how each consumer makes decisions in managing home appliances. In the context of the knapsack problem, each consumer needs to make the best choice to use certain appliances based on the available cost budget and their daily needs. This involves considering basic needs and the desire to use more advanced appliances that use more energy.

Table 3: List of available appliances based on three different cases

Appliances Parameter (i)	**Description of Appliances**	**Case 1**	**Case 2**	**Case 3**
1	Refrigerator	✓	✓	✓
2	Fan	✓	✓	✓
3	Lamp	✓	✓	✓
4	Charger Device	✓	✓	✓
5	Washing Machine	✓	✓	✓
6	Iron	✓	✓	✓
7	Rice Cooker	✓	✓	✓
8	Television	✓	✓	✓
9	Freezer	✓	✓	✓
10	Air Fryer	✓	✓	✓
11	Vacuum		✓	✓
12	Water Purifier	✓		
13	Wi-Fi		✓	✓
14	Pressure Cooker	✓	✓	✓
15	Air Conditioning		✓	✓
16	Microwave			

In this study, CPLEX was used to maximize the EC of the selected appliances without exceeding the cost budget. CPLEX was used in this study because it can quickly and efficiently suggest the duration of usage for the appliances with minimal user involvement [15]. Below is the implementation and the constraints listed for each variable for three different cases with a cost budget, is RM4.75 and RM5.00. RM4.75 per day because based on [8], each operation hour is set as 1kWh. For one day, it is equal to 24kWh equivalent to RM4.75. RM5.00 per day because this study rounded the budget cost of RM4.75 to the nearest Ringgit Malaysia (RM). After

inserting the constraints into the CPLEX and running the coding process, the software will suggest the optimal duration of usage of each appliance per day that can be used without exceeding the cost budget. The results from these two cost budgets will be compared to see the difference in the suggested usage time. This equation is based on (2).

$$\begin{aligned}\text{Maximize} = {} & 0.400x_1 + 0.054x_2 + 0.030x_3 + 0.005x_4 + 0.425x_5 + 1.000x_6 + \\ & 0.300x_7 + 0.100x_8 0.140x_9 + 1.400x_{10} + 0.600x_{11} + 0.300x_{12} + \\ & 0.006x_{13} + 1.000x_{14} + 0.750x_{15} + 0.700x_{16}\end{aligned}$$

Subject to:

$$\begin{aligned}& 0.0792x_1 + 0.0107x_2 + 0.0059x_3 + 0.0010x_4 + \\ & 0.0842x_5 + 0.1980x_6 + 0.0594x_7 + 0.0198x_8 + \\ & 0.0277x_9 + 0.2772x_{10} + 0.1188x_{11} + 0.0594x_{12} + \\ & 0.0012x_{13} + 0.1980x_{14} + 0.1485x_{15} + 0.1386x_{16} \le W\end{aligned}$$

; Cost of appliances for each hour

Table 4 is the list of constraints based on the available appliances for three different cases. Compulsory appliances needed means everyone has the appliances. All three cases use the same constraints. However, there are different constraints for some appliances because of the availability of the appliances. If the consumers did not have the appliances, it would equal to 0. Range of the duration of usage for the appliances based on the average respondents use the appliances from the questionnaire.

Table 4: List of constraints based on the available appliances for three different cases

	Compulsory appliances needed	**Appliances to use when necessary**	**Do not have the appliances**
Case 1	$x_1, x_9, x_{12} = 24$ $8 \le x_2 \le 24$ $5 \le x_3 \le 24$ $1 \le x_4, x_8 \le 4$	$0 \le x_{10}, x_{14} \le 2$	$x_{11}, x_{13}, x_{15}, x_{16} = 0$
Case 2		$0 \le x_{10}, x_{14} \le 2$	$x_{12}, x_{16} = 0$
Case 3		$1 \le x_5, x_6, x_7, x_{11} \le 2$	$x_{10}, x_{13}, x_{14}, x_{15}, x_{16} = 0$

Performance Evaluation

This study will compare before and after using the knapsack model for the total cost consumed by the appliances per day. Real-time monitoring is used to produce comprehensive assessments. The total cost consumed by the appliances per day can be calculated by (3). The cost for each time slot, C_i is RM 0.198 after adjusting the tariff rate and Imbalance Cost Pass-Through (ICPT) rebate.

$$C_T = \sum_{i=1}^{n} P_i \, x_i \, y_i \, C_i \tag{3}$$

where,

C_T	=	Total cost consumed by the appliances per day
i	=	Appliance
n	=	The total number of appliances
P_i	=	The power rating of the appliances i (kilowatts)
x_i	=	The duration of usage for the appliances i per day (hours)
y_i	=	The decision variable for appliance i (1 if selected, 0 otherwise)
C_i	=	The cost for each time slot (RM/kWh)

Total cost for case 1,

$$\begin{aligned} C_T = {} & P_1x_1y_1C_1 + P_2x_2y_2C_2 + P_3x_3y_3C_3 + P_4x_4y_4C_4 + P_5x_5y_5C_5 + P_6x_6y_6C_6 + \\ & P_7x_7y_7C_7 + P_8x_8y_8C_8 + P_9x_9y_9C_9 + P_{10}x_{10}y_{10}C_{10} + P_{11}x_{11}y_{11}C_{11} + \\ & P_{12}x_{12}y_{12}C_{12} + P_{13}x_{13}y_{13}C_{13} + P_{14}x_{14}y_{14}C_{14} + P_{15}x_{15}y_{15}C_{15} + P_{16}x_{16}y_{16}C_{16} \end{aligned}$$

$$\begin{aligned} = {} & 0.400(24)(1)(0.198) + 0.054(24)(1)(0.198) + 0.030(12)(1)(0.198) + \\ & 0.005(3)(1)(0.198) + 0.425(2)(1)(0.198) + (1.000)(1)(1)(0.198) + \\ & 0.300(1)(1)(0.198) + 0.100(5)(1)(0.198) + 0.140(24)(1)(0.198) + \\ & 1.400(1)(1)(0.198) + 0.600(1)(1)(0.198) + 0.300(24)(0)(0.198) + \\ & 0.006(24)(1)(0.198) + 1.000(1)(1)(0.198) + 0.750(8)(1)(0.198) + \\ & 0.700(1)(0)(0.198) \end{aligned}$$

$$= \text{RM } 5.32$$

Total cost for case 2,

$$\begin{aligned} C_T = {} & P_1x_1y_1C_1 + P_2x_2y_2C_2 + P_3x_3y_3C_3 + P_4x_4y_4C_4 + P_5x_5y_5C_5 + P_6x_6y_6C_6 + \\ & P_7x_7y_7C_7 + P_8x_8y_8C_8 + P_9x_9y_9C_9 + P_{10}x_{10}y_{10}C_{10} + P_{11}x_{11}y_{11}C_{11} + \\ & P_{12}x_{12}y_{12}C_{12} + P_{13}x_{13}y_{13}C_{13} + P_{14}x_{14}y_{14}C_{14} + P_{15}x_{15}y_{15}C_{15} + P_{16}x_{16}y_{16}C_{16} \end{aligned}$$

$$= 0.400(24)(1)(0.198) + 0.054(24)(1)(0.198) + 0.030(12)(1)(0.198) + 0.005(3)(1)(0.198) + 0.425(2)(1)(0.198) + (1.000)(1)(1)(0.198) + 0.300(1)(1)(0.198) + 0.100(5)(1)(0.198) + 0.140(24)(1)(0.198) + 1.400(1)(1)(0.198) + 0.600(1)(1)(0.198) + 0.300(24)(0)(0.198) + 0.006(24)(1)(0.198) + 1.000(1)(1)(0.198) + 0.750(8)(1)(0.198) + 0.700(1)(0)(0.198)$$

$$= \text{RM } 5.23$$

Total cost for case 3,

$$C_T = P_1x_1y_1C_1 + P_2x_2y_2C_2 + P_3x_3y_3C_3 + P_4x_4y_4C_4 + P_5x_5y_5C_5 + P_6x_6y_6C_6 + P_7x_7y_7C_7 + P_8x_8y_8C_8 + P_9x_9y_9C_9 + P_{10}x_{10}y_{10}C_{10} + P_{11}x_{11}y_{11}C_{11} + P_{12}x_{12}y_{12}C_{12} + P_{13}x_{13}y_{13}C_{13} + P_{14}x_{14}y_{14}C_{14} + P_{15}x_{15}y_{15}C_{15} + P_{16}x_{16}y_{16}C_{16}$$

$$= 0.400(24)(1)(0.198) + 0.054(24)(1)(0.198) + 0.030(12)(1)(0.198) + 0.005(3)(1)(0.198) + 0.425(2)(1)(0.198) + (1.000)(1)(1)(0.198) + 0.300(1)(1)(0.198) + 0.100(5)(1)(0.198) + 0.140(24)(1)(0.198) + 1.400(1)(0)(0.198) + 0.600(1)(1)(0.198) + 0.300(24)(1)(0.198) + 0.006(24)(1)(0.198) + 1.000(1)(0)(0.198) + 0.750(8)(0)(0.198) + 0.700(1)(1)(0.198)$$

$$= \text{RM } 5.13$$

Results and Discussions

The results of this study provide an analysis of the duration of usage for each appliance using the knapsack model. By using energy consumption (EC) as the value of the appliances and the cost of appliances as the weight, this study gets the optimal solution for the duration of usage for each appliance without exceeding the cost budget which is RM4.75 and RM5.00. The objective is to choose the combination of appliances that stay within the set budget while maximizing EC. For all cases, the optimal duration of usage of appliances that can stay within the cost budget is shown in Table 5. It shows the comparison between the actual total cost of electricity per day and the suggested total cost of electricity per day after using the knapsack model.

Table 5: Summary of the results from CPLEX for the cost budget

		Cost Budget (RM)			
		4.75		5.00	
	Actual Cost	Suggested Cost	Saving Cost	Suggested Cost	Saving Cost
Case 1	5.32	4.66	0.66	4.91	0.41
Case 2	5.23	4.66	0.57	4.91	0.32
Case 3	5.13	4.66	0.47	4.91	0.22

For all cases, the suggested total cost per day is RM4.66, which is less than the cost budget of RM4.75 while RM4.91 for the cost budget of RM5.00 per day. The cost savings range is from RM0.22 to RM0.66 per day and RM6.60 to RM19.80 per month. It shows that this model can help control the monthly electricity bills by ensuring daily EC is within a set budget range. This model also maximizes energy efficiency by prioritizing appliances, such as lights and fans, over less efficient appliances such as air fryers. This not only saves costs but also increases consumer awareness of responsible EC. This can make consumers more enjoyable with a sustainable and budget-friendly lifestyle. The results of CPLEX for each case are shown in Table 6. From the table, it shows that the knapsack model successfully identifies the most energy-efficient appliances within the cost constraints. An important insight is the adaptability of the model to different cost constraints. When the budget increases, the model includes more high EC appliances such as fan and lamp. On the other hand, selecting cheaper and energy-efficient appliances is preferred by those with smaller budgets. This flexibility shows how the knapsack model can be used to address a variety of home energy management (HEM) needs.

Table 6: Summary of the results from CPLEX for the duration of usage for the appliances per day

		Cost Budget (RM)					
		4.75			**5.00**		
		Duration of Usage for the Appliances per Day (hours)					
Appliances Parameter (i)	**Description of Appliances**	**Case 1**	**Case 2**	**Case 3**	**Case 1**	**Case 2**	**Case 3**
1	Refrigerator	24	24	24	24	24	24
2	Fan	22	10	10	23	22	22
3	Lamp	5	5	5	16	21	5
4	Charger Device	3	4	4	1	3	2
5	Washing Machine	1	1	1	1	1	2
6	Iron	1	1	1	1	1	1
7	Rice Cooker	1	2	1	1	1	1
8	Television	3	1	2	2	1	4
9	Freezer	24	24	24	24	24	24
10	Air Fryer	0	0	0	0	0	0
11	Vacuum	0	1	1	0	1	1
12	Water Purifier	24	0	24	24	0	24
13	Wi-Fi	0	24	24	0	24	24
14	Pressure Cooker	0	1	0	0	0	0
15	Air Conditioning	0	8	0	0	8	0
16	Microwave	0	0	0	0	0	0

Conclusion and Recommendations

The knapsack model helps to get efficient energy resource allocation based on consumer preferences and the energy efficiency of appliances. The first objective of this study was to adapt an optimization usage of appliances per day using the knapsack model for home energy management (HEM) of home appliances. The focus of home appliances is to optimize energy efficiency and balance the household energy load. Some appliances such as refrigerator, water purifier, freezer and wi-fi must be used continuously which is 24 hours and always prioritized while other appliances are scheduled based on the capacity limit which is the cost budget of RM4.75 and RM5.00. Appliances such as air fryer, pressure cooker and microwave will be selected if the total cost consumed by appliances does not exceed the cost budget.

The second objective is to compare the cost of electricity bills per day before and after using the knapsack model. The findings show that strategically scheduling appliance usage with the model enhances energy efficiency, potentially reducing electricity costs. This analysis shows the significance of optimization in household energy management and emphasizes the benefits of efficient appliance scheduling. Overall, the results support the knapsack model as a valuable tool for reducing electricity costs while ensuring essential energy needs are met.

The knapsack model is one of the effective models to solve problems related to optimization. Future studies can include more appliances with varying energy demands to improve the knapsack model's performance. This addition will allow for more accurate and efficient management overall energy consumption (EC). Besides, renewable energy such as solar panels may be considered in the optimization process. The knapsack model may produce specific energy management solutions that encourage increased efficiency and support sustainable energy practices by adjusting to both conventional and renewable energy systems, especially in different weather conditions

Acknowledgments

The author would like to express their heartfelt appreciation to everyone who contributed to this study, with special attention to the family, supervisor, all the respondents, and friends for their dedication, efforts, invaluable guidance and support throughout this journey.

References

[1]Shivanaganna, N., & Murali, V. (2022). Energy routing based optimal solution for residential peak power reduction with photovoltaic and electric vehicle penetration. Energy Sources. Part A, Recovery, Utilization, and Environmental Effects, 44(3), 5682–5698. https://doi.org/10.1080/15567036.2022.2085824

[2]Qayyum, N., Amin, A., Jamil, U., & Mahmood, A. (2019). Optimization Techniques for Home Energy Management: A Review. 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). https://doi.org/10.1109/icomet.2019.8673435

[3]Hischier, R., Reale, F., Castellani, V., & Sala, S. (2020). Environmental impacts of household appliances in Europe and scenarios for their impact reduction. Journal of Cleaner Production, 267, 1-12. https://doi.org/10.1016/j.jclepro.2020.121952

[4] Balogun, O. A., Dosunmu, M. A., & I.O Ibidapo. (2022). An Explanatory Comparison of Brute Force and Dynamic Programming Techniques to Solve the 0-1 Knapsack Problem. Zenodo (CERN European Organization for Nuclear Research), 7(2), 660-664. https://doi.org/10.5281/zenodo.6345484.

[5]Wong, J. C. Y., Blankenship, B., Urpelainen, J., Balani, K., Ganesan, K., & Bharadwaj, K. (2022). Understanding electricity billing preferences in rural and urban India: Evidence from a conjoint experiment. Energy Economics, 106, 1-12. https://doi.org/10.1016/j.eneco.2021.105735

[6]Kumabe , S., & Yoshida, Y. (2022). Average Sensitivity of the Knapsack Problem. 244(14), 1–75. https://doi.org/10.4230/LIPIcs.ESA.2022.75

[7] Boes, D. F., Dewanto, K. P., Raushanfikir, M. I., Handoyo, A. T., & Nurhasanah, N. (2023). Analyzing The Most Effective Algorithm For Knapsack Problems. 172-176. https://doi.org/10.1109/icon-sonics59898.2023.10434966

[8] Anzar, M., Iqra, R., Kousar, A., Ejaz, S., Alvarez-Alvarado, M. S., & Zafar, A. K. (2017). Optimization of Home Energy Management System in Smart Grid for Effective Demand Side Management. https://doi.org/10.1109/irsec.2017.8477255

[9]Chen, X. (2022). A Comparison of Greedy Algorithm and Dynamic Programming Algorithm. SHS Web of Conferences, 144, 1-5. https://doi.org/10.1051/shsconf/202214403009

[10]Malik, S. A., & Ayop, A. R. (2020). Solar energy technology: Knowledge, awareness, and acceptance of B40 households in one district of Malaysia towards government initiatives. Technology in Society, 63, 1-7. https://doi.org/10.1016/j.techsoc.2020.101416

[11]Siebra, C., Costa, P., Da Silva, F. Q., Santos, A. L., & Mascaro, A. (2013). The hardware and software aspects of energy consumption in the mobile development platform. International Journal of Pervasive Computing and Communications, 9(2), 139–162. https://doi.org/10.1108/ijpcc-04-2013-0007

[12]Shewale, A., Mokhade, A., Lipare, A., & Bokde, N. D. (2023). Efficient Techniques for Residential Appliances Scheduling in Smart Homes for Energy Management Using Multiple Knapsack Problem. Arabian Journal for Science and Engineering, 49(3), 3793–3813. https://doi.org/10.1007/s13369-023-08178-w

[13]Iliyasu, R., & Etikan, I. (2021). Comparison of Quota Sampling and Stratified Random Sampling. Biometrics & Biostatistics International Journal, 10(1), 24–27. https://doi.org/10.15406/bbij.2021.10.00326

[14] Ali, S. S. S., Razman, M. R., & Awang, A. (2020). The Estimation and Relationship of Domestic Electricity Consumption and Appliances Ownership in Malaysia Intermediate City. International Journal of Energy Economics and Policy, 10(6), 116–122. https://doi.org/10.32479/ijeep.8358

[15]Lu, Y., & Vasko, F. J. (2019). A comprehensive empirical demonstration of the impact of choice constraints on solving generalizations of the 0–1 knapsack problem using the integer programming option of CPLEX®. Engineering Optimization, 52(9), 1632–1644. https://doi.org/10.1080/0305215x.2019.1658748

Prediction Model of Home Loan Eligibility Status

Danisya Adlina Ruzi[1], Siti Aisyah Mohamed Kher[1], Isnewati Ab Malek[1,], Haslinda Ab Malek[1*,]
[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia*
[*]Corresponding: haslinda8311@uitm.edu.my

Abstract

A home loan is known as a mortgage which is a type of loan used to purchase a home or real estate. Home loans are essential for many individuals who are looking to buy a home without paying the entire price upfront. Confronting a prevalent challenge involving a notable of 60% rejection rate among home loan applicants, this issue was identified during the pre-pandemic year. Therefore, this study aims to develop an optimized home loan eligibility status prediction model. The study systematically evaluated Decision Tree approaches, incorporating variable selection for enhanced precision in model comparison. The findings of this study found that loan amount, marital status, property area, and co-applicant income as pivotal factors influencing home loan eligibility status. Notably, the Decision Tree Entropy model emerges as the optimal model, achieving a remarkable average squared error of 19.71%. This model exhibits superior performance, evidenced by a high sensitivity of 84.16% and accuracy of 59.78% according to the confusion matrix. The results are hoped to provide valuable insights for improving house loan eligibility prediction models, which could improve the banking sector decision-making procedures and in turn, improve consumer satisfaction.

Keywords*:* Housing loan, Loan Approval, Decision Tree.

Introduction

A home loan is a loan that helps a borrower purchase a home. The amount borrowed from a home is based on the appraised value of the home and the amount of money paid as a down payment. The home will be put up as collateral and loan applicants will make monthly mortgage payments for a specified amount of time (typically 15 or 30 years) until the loan is paid off. According to a study done by [1], in Malaysia, home loans can be divided into three main categories: Basic Term Loans, Semi-Flexi Loans, and Full-Flexi Loans.

In this study, the home loan eligibility status is predicted. The term 'home loan eligibility status' defines an applicant's suitability for a home loan based on the requirements set by the lender. It indicates if the applicants meet the requirements for a home loan based on variables such as income and loan amount. Home loan eligibility status is an important factor in the lending industry, and financial institutions must predict the likelihood of loan approval accurately. However, according to the finding by [2], an average of 60% of the home loan applicants failed to get a home loan during the pre-pandemic year. This has been a problem and a significant issue for the home loan applicants.

Therefore, this study was conducted to construct the home loan eligibility status model in determining the type of model that is best to use. The model can help to analyse applicant data more rapidly and improve the overall customer experience by offering faster and more accurate predictions. The application of these models can assist banks and financial institutions to make more informed home loan approval choices. Moreover, these models can reduce the risk of loan default and improve the overall efficiency of the home loan approval process.

As the demand for home loans in banks grows by the day, [3] performed a Modern Approach to Home Loan Sanctioning in Banks Using Machine Learning. The goal is to apply machine learning tools that use classification algorithms to predict worthy applicants for home loan approval as an appropriate option to reduce human efforts and make successful decisions in the home loan approval process. A system to develop a model is built in their research by training the system with records and approval results of previously applied home loan applicants

Moreover, study done by [4] employed two machine learning classification models, namely Decision Tree and Random Forest, for loan prediction. The three findings indicate that the highest level of accuracy achieved is

62.12%, accompanied with a confidence factor of 0.15. Other than that, [5] employed decision trees to construct an accurate and efficient predictive model for forecasting home loan approval. Experiments were conducted utilising several tree methodologies, spanning from the simplest and most comprehensible decision tree to the most intricate random forests.

Hence, [6] did a comprehensive study to make a detailed comparison between the Random Forest and Decision Trees algorithms. Both methodologies were employed on the identical dataset, and the outcomes revealed that the Random Forest algorithm exhibited much superior performance in terms of accuracy compared to the Decision Tree strategy, achieving a score of 80% as opposed to 73%. Moreover, [7] conducted a study to forecast the eligibility of consumers for home loan approval. Three machine learning models employed were the Logistic Regression Model, the Decision Tree, and the Random Forest. The study revealed that based on the comparison of the three models, Logistic Regression exhibits the highest accuracy rate of 89.7059%, followed by Decision Tree with an accuracy rate of 85.4054% and Random Forest with an accuracy rate of 77.4566%.

Based on the feature analysis run by [8], it showed that those who have a higher credit history have a better probability of getting a loan compared to those who do not. Furthermore, [9] conducted a study stated that the home loan approval status is determined to be independent of the applicant's income, gender, and loan amount. However, it appears that the credit history has the greatest impact on loan acceptance, suggesting that an applicant's credit history is used as an initial filter for loan applications.

The findings of this study will provide banking institutions with the best model for determining eligibility for housing loans. The greater demand for property loans justifies the need for a more effective, speedy, and prompt approach for bankers in predicting potential candidates. For the researchers, the study will help them uncover the best model for predicting the loan approval process that many researchers were not able to explore. Moreover, this study not only helps the applicant but also helps the bank institution by minimizing the risk and reducing the number of defaulters.

Methodology

Source of data

This study employed secondary data collected from GitHub. The dataset originally sourced from Analytics Vidhya's Loan Prediction Challenge, comprised 443 observations and 11 variables, with Loan Status serving as the dependent variable of interest [10]. The variables used in the analysis are summarized in Table 1.

Table 1: Description of the Variables

No	Variable	Type of Variables	Scale	Description
1	Gender	Categorical	Nominal	Male, Female
2	Marital Status	Categorical	Nominal	Yes, No
3	Number of Dependents	Continuous	Interval	0, 1, 2, 3+
4	Education Status	Categorical	Nominal	Graduate, Not Graduate
5	Self-Employment	Categorical	Nominal	Yes, No
6	Applicant's Income	Continuous	Interval	$0 - $22000
7	Co-Applicant's Income	Continuous	Interval	$0 - $11000
8	Loan Amount	Continuous	Interval	$15000 - $375000
9	Loan Term	Continuous	Interval	126.68 - 480 (months)
10	Property Area	Categorical	Nominal	Urban, Semiurban, Rural
11	Loan Status	Categorical	Nominal	Yes, No

Decision Tree

The purpose of employing a Decision Tree in this study is to develop a model that can be used to predict the class or value of the home loan eligibility status by learning simple decision rules inferred from prior data. This study employed three of the most common types of Decision Trees, namely Gini, CHAID, and Entropy. These types are distinguished from each other by the mathematical model used to select splitting property when extracting Decision Tree rules [11]

Model Assessment

Accuracy measures can be used to evaluate the ability of the estimated model to predict the target. The Receiver Operating Characteristic (ROC) Chart, Confusion Matrix, Misclassification Rate, Average Squared Error (ASE), and Receiver Operating Characteristic (ROC) Index are the tools from which different accuracy measures are derived.

Misclassification Rate

The misclassification rate is the frequency with which a classification model guesses a sample's class or category in a dataset incorrectly. It serves as an indicator of how well the model classifies the data. When assessing a classification model, the misclassification rate is determined by comparing the sample true class labels with the predicted class labels. The misclassification rate, for example, would be 10% if a model properly classified 90 out of 100 samples, meaning that 10% of the data were misclassified.

Average Squared Error

The average squared error (ASE) measures how close a regression line comes to a set of data points. The model with a smaller value of ASE indicates a better fit. Hence, it is better suited to be used in predicting the target. Equation 1, as shown, can be used to calculate the average squared error

$$\mathrm{ASE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{1}$$

Receiver Operating Characteristic (ROC) Index

The Receiver Operating Characteristic (ROC) index is the area under the value (AUC) in the ROC chart. The AUC is widely used to measure the accuracy of diagnostic tests. The higher the value of the ROC index, the greater the sensitivity and accuracy of the model.

Confusion Matrix

Accuracy is measured regarding the error rate percentage of records classified correctly or incorrectly. A confusion matrix can be used to determine this measurement. The number of accurate and inaccurate predictions made by the classification model concerning the actual target variable in the data is displayed in a confusion matrix.

Results and Discussion

Model comparisons between Decision Tree (DT) Gini, Decision Tree (DT) Entropy and Decision Tree (DT) CHAID were computed to compare the training sample and validation sample based on the values of the misclassification rate, average squared error, and Receiver Operating Characteristic (ROC) index.

Table 2: Model Comparison Based on Misclassification Rate, Average Squared Error and ROC Index

Model	Misclassification Rate			Average Squared Error			ROC Index		
	Valid	Train	Gap	Valid	Train	Gap	Valid	Train	Gap
DT Gini	0.430	0.307	0.123	0.278	0.197	0.081	0.494	0.735	-0.241
DT Entropy	0.413	0.303	0.110	0.291	0.197	0.094	0.570	0.722	-0.152
DT CHAID	0.458	0.386	0.072	0.247	0.232	0.015	0.559	0.621	-0.062

Table 2 shows the values of valid, train, and gap for the misclassification rate, average squared error, and ROC index for each model. The valid and gap values are used to find the best model. The gap values are calculated by finding the difference between valid and train values. The negative sign is not taken into consideration for the gap value. From Table 2, it is found that the DT Gini has the highest gap value for the misclassification rate (0.123) and ROC index (0.241). Therefore, DT Gini is found to be the overfit model.

Additionally, the valid value is evaluated to determine the best fit model. The best model has the lowest valid values for the misclassification rate and average squared error. In contrast, the model must have the highest valid value of the ROC index. DT Entropy has the lowest valid misclassification rate (0.413) and the highest ROC index valid value (0.570). Hence, DT Entropy is concluded as the best model since it meets most of the conditions.

Table 3 summarise that there is no underfit model since all model performances have the best validation results and no negative valid value. However, there is one overfit model, which belongs to the Decision Tree Gini. An overfit model can be determined when it has the largest gap value between valid and train value. Thus, the Decision Tree Gini is an overfitted model because it has the largest gap values for the misclassification rate and the ROC index. In choosing the best models, the valid values for the misclassification rate and average squared error need to be the lowest, while the value for the ROC index needs to be the highest. The Decision Tree Entropy meets most of the requirements, hence, it is chosen as the best model.

Table 3: Summary of Model Comparison

Assesment of Generalization	**Model Description**
Underfit	None
Overfit	Decision Tree Gini
Best Fit	Decision Tree Entropy

Best Model

The confusion matrix is observed to further confirm that the Decision Tree Entropy is the best model. A confusion matrix is a table that describes the performance of a classification model on a set of test data. The factors in the confusion matrix can be examined to yield several performance measures such as sensitivity, specificity, and accuracy, which provide insights into the model's performance.

Table 4: Result Analysis of Confusion Matrix

Model	**Sensitivity**	**Specificity**	**Accuracy**
Decision Tree Gini	0.6931	0.4231	0.5754
Decision Tree Entropy	0.8416	0.2821	0.5978
Decision Tree CHAID	0.4752	0.6281	0.5419

Based on Table 4, the Decision Tree CHAID has the highest specificity values at 62.81%, showing that the model is the best model for classifying the 'No' group of home loan eligibility status. Meanwhile, the Decision Tree Entropy

has the highest sensitivity and accuracy values at 84.16% and 59.78%, respectively. This confirmed that Decision Tree Entropy is the best model for classifying the 'Yes' group of home loan eligibility status and the best in predicting the overall performance of the model.

Conclusion

The specific requirements for obtaining a home can be ambiguous, causing applicants to question the qualifying conditions and the procedures involved in applying. Thus, this study has been conducted using secondary data obtained which includes ten variables that are believed to have an impact on the approval of house loans. This dataset has 443 observations with a binary target variable. The Decision Tree Gini is found to be overfit while none of the models are underfit. It was discovered that implementing Decision Tree Entropy produces the best model for predicting a house loan applicant's eligibility status with an average squared error of only 19.71% and highest accuracy values of 59.78%. Consequently, the Decision Tree Entropy model was deemed to be the most reliable in predicting home loan eligibility status. It is hoped that this research provides valuable insights for home-loan applicants, banking institutions and other researchers for their own interest. This study is in line with a study done [12], it indicates the entropy reduction decision tree technique outperforms the other seven methods in terms of overall classification accuracy rates at the 0.5 cut-off.

For future studies, it is recommended explore other models apart from DT Gini, Decision Tree Entropy and Decision Tree CHAID. One example of an alternative decision tree model is the Random Forest algorithm. Random Forest combines multiple decision trees to make predictions, offering improved accuracy and reducing the risk of overfitting. By utilising a variety of models, researchers can evaluate their performance and compare the results to identify the most effective model for predicting home loan eligibility status.

References

[1] Hassan, M. M., Ahmad, N., & Hashim, A. H. (2022). Housing Property Investment Opportunity in Malaysia: Things that new investor should know. International Journal of Academic Research in Business and Social Sciences, 12(1), 924–941. https://doi.org/10.6007/IJARBSS/v12-i1/12003.

[2] Sooi, C. C. (2021, Sep). Mitigating malaysia's worrying trend of high home loan rejection focus malaysia. Retrieved from https:// focusmalaysia.my/mitigating-malaysias-worrying-trend-of-high-home-loan-rejection.

[3] Rath, G. B., Das, D., & Acharya, B. (2021). Modern approach for loan sanctioning in banks using machine learning. In Advances in machine learning and computational intelligence: Proceedings of icmlci 2019 (pp. 179–188).

[4] Gautam, et al. (2020). Loan prediction using decision tree and random forest. International Research Journal of Engineering and Technology (IRJET), 7(08), 853–856.

[5] Alaradi, M., & Hilal, S. (2020). Tree-based methods for loan approval. In 2020 international conference on data analytics for business and industry: Way towards a sustainable economy (icdabi) (pp. 1–6).

[6] Madaan, et al. (2021). Loan default prediction using decision trees and random forest: A comparative study. In Iop conference series: Materials science and engineering (Vol. 1022, p. 012042).

[7] Dutta, P. (2021). A study on machine learning algorithm for enhancement of loan prediction. International Research Journal of Modernization in Engineering Technology and Science, 3.

[8] Sunitha, T., Chandravallika, M., Ranganayak, M., Suma Sri, G., Jagadeesh, T. V. S., & Tejaswi, A. (2021). Predicting the Loan Status using Logistic Regression and Binary Tree. Paper presented at ICICNIS 2020. http://dx.doi.org/10.2139/ssrn.3769854.

[9] Alaradi, M., & Hilal, S. (2020). Tree-based methods for loan approval. In 2020 international conference on data analytics for business and industry: Way towards a sustainable economy (icdabi) (pp. 1–6).

[10] Temburwar, S. (2019). Loan Prediction Dataset [Data set]. GitHub. https://github.com/shrikant-temburwar/Loan-Prediction-Dataset.

[11] Frempong, N. K., Nicholas, N., & Boateng, M. (2017). Decision tree as a predictive modeling tool for auto insurance claims. International Journal of Statistics and Applications, 7(2), 117–120

[12] Zurada, J. (2010). Could decision trees improve the classification accuracy and interpretability of loan granting decisions? In 2010 43rd hawaii international conference on system sciences (p. 1-9). doi: 10.1109/HICSS.20

RADIUS OF STARLIKENESS OF A SUBCLASS OF GENERALIZED BI-UNIVALENT FUNCTIONS

Nur Suhailah Binti Zamri[1], Abdullah Yahya[1,*]
[1]*College of Computing, Informatics and Mathematics,*
UiTM Cawangan Negeri Sembilan, Kampus Seremban, Malaysia.
[*]abdullahyahya@uitm.edu.my

Abstract

In this paper, we let a Taylor series expansion, $S^*(\alpha,\delta,t)$ *of the form* $g(m)=m+a_2m^2+a_3m^3+\ldots+\sum_{z=2}^{\infty}a_zm^z$ *that satisfies the condition* $Re\left(e^{i\alpha}\frac{mg'(m)}{g'(m)}\right)>\delta,\ m\in E$ *where* $|\alpha|<\pi,\ \cos\alpha>\delta,\ 0\le\delta<1,\ -1<t\le 1$ *and* $g'(m)=\frac{m}{(1+tm)(1-m)}$. *The main focus of this article is to determine the radius of starlikeness,* $\Re_{St}$ *by using Poisson Formula for the new generalized class of bi-univalent functions,* $S^*(\alpha,\delta,t)$.

Keywords*:* Radius of starlikness, Bi-univalent, Poisson formula

Introduction

Complex analysis is one of the properties in mathematics that deal with functions that involve complex numbers, including their functions, manipulation, derivatives and other mathematical properties. The geometric function theory is a study of the geometric properties of univalent functions. This paper focuses on the class of analytic univalent functions implied with the complex plane. Let $\mathbb{C}$ be the element of complex number and let $g(m)$ be a complex valued function of the complex variable m. Based on Goodman [1], a function $g(m)$ is univalent in $D\subset\mathbb{C}$ if the function $g(m)$ is injective, which is for all $m_1,m_2\in D$, $g(m_1)=g(m_2)$. Besides, a function $g(m)$ provide one to one mapping onto its image, $g(D)$.

Let Z denoted the class of function of the form

$$g(m)=m+a_2m^2+a_3m^3+\ldots+\sum_{z=2}^{\infty}a_zm^z \tag{1.1}$$

which are analytic in the unit disc, $E=\{m:|m|<1\}$. The function $g(m)$ is also known as normalized univalent function if it satisfies conditions of $g(0)=0$ and $g'(0)=1$ or $g(x)=g'(0)-1=0$ are fixed is denoted by S (Duren, [2])

In this paper, we defined a new generalized class of bi-univalent function $S^*(\alpha,\delta,t)\in S$. In this class, $S^*(\alpha,\delta,t)$ has a Taylor series expansion of the form (1.1) that satisfying the condition

$$Re\left(e^{i\alpha}\frac{mg'(m)}{g'(m)}\right)>\delta \quad (m\in E)$$

where $|\alpha|<\pi,\ \cos\alpha>\delta,\ 0\le\delta<1,\ g'(m)=\dfrac{m}{(1+tm)(1-m)}$ and $-1<t\le 1$.

Based on $S^*(\alpha,\delta,t)$, the coefficient bound is given following lemma derived from Rathi [3] and the method of proving by Nunokawa and Owa [4].

Lemma 1.1

Suppose that $f\in S^*(\alpha,\delta,t)$ is given by $(1.1.1)$, then sharp inequality

$$|a_n|\le\begin{cases}\dfrac{1}{n}\left(\dfrac{1-t^2+2A_{\alpha\delta}\left[(-1+t)+n(t+1)\right]}{(t+1)^2}\right), & n=2,4,6,\ldots\\ \dfrac{1}{n}\left(\dfrac{2A_{\alpha\delta}(n-1)+t+1}{(t+1)}\right), & n=3,5,7,\ldots\end{cases}$$

equality is attained for each n while f is extreme point of $S^*(\alpha,\delta,t)$.

Lemma 1.2 (Ikeda and Saigo, [5])

Let $u(m)$ be harmonic in $|m|\le\rho$ and continuous in $|m|\le\rho$. Then $u(m)$ is given by the equation

$$u(m)=u(re^{i\theta})=\frac{1}{2\pi}\int \mathrm{Re}\{u(\rho e^{i\phi})\}\frac{\rho^2-r^2}{\rho^2-2\rho r\cos(\phi-\theta)+r^2}d\phi \qquad (1.2)$$

where $0\le r\le\rho$

The equation (1.2) is called the Poisson formula. Thus, by following the method of Nunokawa and Owa [4], the radius of starlikeness will be determined using this formula.

Preliminaries

Based on Lemma 1.1, we are going to find a bound for $\left|\frac{g(m)}{m}\right|$ by using the $|a_n| = \frac{1}{n}\left(\frac{1-t^2+2A_{\alpha\delta}\left[(-1+t)+n(t+1)\right]}{(t+1)^2}\right)$ when n is even.

Theorem 2.1

Let $g \in Z$, *since* $g \in S^*(\alpha,\delta,t)$ *and* $|a_n| = \frac{1}{n}\left(\frac{1-t^2+2A_{\alpha\delta}\left[(-1+t)+n(t+1)\right]}{(t+1)^2}\right)$ *where* $n = 2,4,6,\ldots,$ *then*

$$\log(1-r)\left[\frac{-t^2+2tA_{\alpha\delta}-2A_{\alpha\delta}+1}{r(t+1)^2}\right]+\left[1+\left(\frac{-t^2+4tA_{\alpha\delta}+1}{(t+1)^2}\right)\right]-\left(\frac{1}{1-r}\right)\left[\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{r(t+1)^2}\right]$$
$$+\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{r(t+1)^2} \le \left|\frac{g(m)}{m}\right| \le \log(1-r)\left[\frac{t^2-2tA_{\alpha\delta}+2A_{\alpha\delta}-1}{r(t+1)^2}\right]+\left[1+\frac{t^2-4tA_{\alpha\delta}-1}{(t+1)^2}\right]$$
$$+\left(\frac{1}{1-r}\right)\left[\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{r(t+1)^2}\right]-\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{r(t+1)^2}$$

for $|m| = r < 1.$

Proof.

Given $g \in Z$, since $g \in \mathbb{N}$ satisfying Lemma 4.1.1 for $n = 2,4,6,\ldots$, we have

$$|g(m)| \le |m| + \sum_{n=2}^{\infty} |a_n||m|^n$$

and

$$|g(m)| \le |m| + \sum_{n=2}^{\infty}\left|\frac{1}{n}\left(\frac{1-t^2+2A_{\alpha\delta}\left[(-1+t)+n(t+1)\right]}{(t+1)^2}\right)\right||m|^n \tag{2.1}$$

and the expansion of (2.1),

$$|g(m)| \le |m| + \sum_{n=2}^{\infty}\left|\frac{1}{n}\left(\frac{1}{(t+1)^2}-\frac{t^2}{(t+1)^2}-\frac{2A_{\alpha\delta}}{(t+1)^2}+\frac{2tA_{\alpha\delta}}{(t+1)^2}+\frac{2tnA_{\alpha\delta}}{(t+1)^2}+\frac{2nA_{\alpha\delta}}{(t+1)^2}\right)\right||m|^n$$

produces

$$|g(m)| \le |m| + \frac{1}{(t+1)^2}\sum_{n=2}^{\infty}\frac{|m|^n}{n} - \frac{t^2}{(t+1)^2}\sum_{n=2}^{\infty}\frac{|m|^n}{n} - \frac{2A_{\alpha\delta}}{(t+1)^2}\sum_{n=2}^{\infty}\frac{|m|^n}{n}$$

$$+ \frac{2tA_{\alpha\delta}}{(t+1)^2}\sum_{n=2}^{\infty}\frac{|m|^n}{n} + \frac{2tA_{\alpha\delta}}{(t+1)^2}\sum_{n=2}^{\infty}|m|^n + \frac{2A_{\alpha\delta}}{(t+1)^2}\sum_{n=2}^{\infty}|m|^n. \quad (2.2)$$

Then, simplification of (2.2), yield

$$|g(m)| \le |m| + \frac{1}{(t+1)^2}\left(-|m| + \sum_{n=1}^{\infty}\frac{|m|^n}{n}\right) - \frac{t^2}{(t+1)^2}\left(-|m| + \sum_{n=1}^{\infty}\frac{|m|^n}{n}\right)$$
$$- \frac{2A_{\alpha\delta}}{(t+1)^2}\left(-|m| + \sum_{n=1}^{\infty}\frac{|m|^n}{n}\right) + \frac{2tA_{\alpha\delta}}{(t+1)^2}\left(-|m| + \sum_{n=1}^{\infty}\frac{|m|^n}{n}\right)$$
$$+ \frac{2tA_{\alpha\delta}}{(t+1)^2}\left(-1-|m| + \sum_{n=0}^{\infty}|m|^n\right) + \frac{2A_{\alpha\delta}}{(t+1)^2}\left(-1-|m| + \sum_{n=0}^{\infty}|m|^n\right).$$

Since $\log(1-|m|) = -\sum_{n=1}^{\infty}\frac{|m|^n}{n}$ and $\frac{1}{1-|m|} = \sum_{n=0}^{\infty}|m|^n$, thus

$$|g(m)| \le |m| + \frac{1}{(t+1)^2}\left(-|m| - \log(1-|m|)\right) - \frac{t^2}{(t+1)^2}\left(-|m| - \log(1-|m|)\right)$$
$$- \frac{2A_{\alpha\delta}}{(t+1)^2}\left(-|m| - \log(1-|m|)\right) + \frac{2tA_{\alpha\delta}}{(t+1)^2}\left(-|m| - \log(1-|m|)\right)$$
$$+ \frac{2tA_{\alpha\delta}}{(t+1)^2}\left(-1-|m| + \frac{1}{1-|m|}\right) + \frac{2A_{\alpha\delta}}{(t+1)^2}\left(-1-|m| + \frac{1}{1-|m|}\right).$$

Which gives

$$|g(m)| \le |m| - \frac{|m|}{(t+1)^2} - \frac{1}{(t+1)^2}\log(1-|m|) + \frac{t^2|m|}{(t+1)^2} + \frac{t^2}{(t+1)^2}\log(1-|m|)$$
$$+ \frac{2A_{\alpha\delta}|m|}{(t+1)^2} + \frac{2A_{\alpha\delta}}{(t+1)^2}\log(1-|m|) - \frac{2tA_{\alpha\delta}|m|}{(t+1)^2} - \frac{2tA_{\alpha\delta}}{(t+1)^2}\log(1-|m|) - \frac{2tA_{\alpha\delta}}{(t+1)^2}$$
$$- \frac{2tA_{\alpha\delta}|m|}{(t+1)^2} + \frac{2tA_{\alpha\delta}}{(t+1)^2}\left(\frac{1}{1-|m|}\right) - \frac{2A_{\alpha\delta}}{(t+1)^2} - \frac{2A_{\alpha\delta}|m|}{(t+1)^2} + \frac{2A_{\alpha\delta}}{(t+1)^2}\left(\frac{1}{1-|m|}\right). \quad (2.3)$$

Next, substitute $|m|=r$ into equation (2.3) as

$$|g(m)| \le \log(1-r)\left[\frac{t^2-2tA_{\alpha\delta}+2A_{\alpha\delta}-1}{(t+1)^2}\right]+r\left[1+\frac{t^2-4tA_{\alpha\delta}-1}{(t+1)^2}\right] + \left(\frac{1}{1-r}\right)\left[\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{(t+1)^2}\right]-\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{(t+1)^2}. \quad (2.4)$$

Thus, multiply equation (2.4) with $\frac{1}{r}$ to obtain the equation for $\left|\frac{g(m)}{m}\right|$,

$$\left|\frac{g(m)}{m}\right| \le \log(1-r)\left[\frac{t^2-2tA_{\alpha\delta}+2A_{\alpha\delta}-1}{r(t+1)^2}\right]+\left[1+\frac{t^2-4tA_{\alpha\delta}-1}{(t+1)^2}\right] + \left(\frac{1}{1-r}\right)\left[\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{r(t+1)^2}\right]-\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{r(t+1)^2}$$

for $|m|=r<1$.

Hence, $g(m)$ absolutely converge in E, also $g(m)$ is analytic in E. Therefore, we have

$$|g(m)| \ge |m|-\sum_{n=2}^{\infty}|a_n||m|^n,$$

$$|g(m)| \ge |m|-\sum_{n=2}^{\infty}\left|\frac{1}{n}\left(\frac{1-t^2+2A_{\alpha\delta}\left[(-1+t)+n(t+1)\right]}{(t+1)^2}\right)\right||m|^n, \quad (2.5)$$

and the expansion of (2.5),

$$|g(m)| \ge |m|-\sum_{n=2}^{\infty}\left|\frac{1}{n}\left(\frac{1}{(t+1)^2}-\frac{t^2}{(t+1)^2}-\frac{2A_{\alpha\delta}}{(t+1)^2}+\frac{2tA_{\alpha\delta}}{(t+1)^2}+\frac{2tnA_{\alpha\delta}}{(t+1)^2}+\frac{2nA_{\alpha\delta}}{(t+1)^2}\right)\right||m|^n$$

produces

$$|g(m)| \ge |m|-\frac{1}{(t+1)^2}\sum_{n=2}^{\infty}\frac{|m|^n}{n}+\frac{t^2}{(t+1)^2}\sum_{n=2}^{\infty}\frac{|m|^n}{n}+\frac{2A_{\alpha\delta}}{(t+1)^2}\sum_{n=2}^{\infty}\frac{|m|^n}{n} - \frac{2tA_{\alpha\delta}}{(t+1)^2}\sum_{n=2}^{\infty}\frac{|m|^n}{n}-\frac{2tA_{\alpha\delta}}{(t+1)^2}\sum_{n=2}^{\infty}|m|^n-\frac{2A_{\alpha\delta}}{(t+1)^2}\sum_{n=2}^{\infty}|m|^n. \quad (2.6)$$

Then, simplification of (2.6) yields

$$|g(m)| \ge |m|-\frac{1}{(t+1)^2}\left(-|m|+\sum_{n=1}^{\infty}\frac{|m|^n}{n}\right)+\frac{t^2}{(t+1)^2}\left(-|m|+\sum_{n=1}^{\infty}\frac{|m|^n}{n}\right)$$

$$+\frac{2A_{\alpha\delta}}{(t+1)^2}\left(-|m|+\sum_{n=1}^{\infty}\frac{|m|^n}{n}\right)-\frac{2tA_{\alpha\delta}}{(t+1)^2}\left(-|m|+\sum_{n=1}^{\infty}\frac{|m|^n}{n}\right)$$
$$-\frac{2tA_{\alpha\delta}}{(t+1)^2}\left(-1-|m|+\sum_{n=0}^{\infty}|m|^n\right)-\frac{2A_{\alpha\delta}}{(t+1)^2}\left(-1-|m|+\sum_{n=0}^{\infty}|m|^n\right).$$

Since $\log(1-|m|)=-\sum_{n=1}^{\infty}\frac{|m|^n}{n}$ and $\frac{1}{1-|m|}=\sum_{n=0}^{\infty}|m|^n$, thus

$$|g(m)|\geq|m|-\frac{1}{(t+1)^2}\left(-|m|-log(1-|m|)\right)+\frac{t^2}{(t+1)^2}\left(-|m|-log(1-|m|)\right)$$
$$+\frac{2A_{\alpha\delta}}{(t+1)^2}\left(-|m|-log(1-|m|)\right)-\frac{2tA_{\alpha\delta}}{(t+1)^2}\left(-|m|-log(1-|m|)\right)$$
$$-\frac{2tA_{\alpha\delta}}{(t+1)^2}\left(-1-|m|+\frac{1}{1-|m|}\right)-\frac{2A_{\alpha\delta}}{(t+1)^2}\left(-1-|m|+\frac{1}{1-|m|}\right).$$

Which gives

$$|g(m)|\geq|m|+\frac{|m|}{(t+1)^2}+\frac{1}{(t+1)^2}log(1-|m|)-\frac{t^2|m|}{(t+1)^2}$$
$$-\frac{t^2}{(t+1)^2}log(1-|m|)-\frac{2A_{\alpha\delta}|m|}{(t+1)^2}-\frac{2A_{\alpha\delta}}{(t+1)^2}log(1-|m|)$$
$$+\frac{2tA_{\alpha\delta}|m|}{(t+1)^2}+\frac{2tA_{\alpha\delta}}{(t+1)^2}log(1-|m|)+\frac{2tA_{\alpha\delta}}{(t+1)^2}+\frac{2tA_{\alpha\delta}|m|}{(t+1)^2}$$
$$-\frac{2tA_{\alpha\delta}}{(t+1)^2}\left(\frac{1}{1-|m|}\right)+\frac{2A_{\alpha\delta}}{(t+1)^2}+\frac{2A_{\alpha\delta}|m|}{(t+1)^2}-\frac{2A_{\alpha\delta}}{(t+1)^2}\left(\frac{1}{1-|m|}\right)$$

$$|g(m)|\geq\log(1-|m|)\left[\frac{-t^2+2tA_{\alpha\delta}-2A_{\alpha\delta}+1}{(t+1)^2}\right]+|m|\left[1+\left(\frac{-t^2+4tA_{\alpha\delta}+1}{(t+1)^2}\right)\right]$$
$$-\left(\frac{1}{1-|m|}\right)\left[\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{(t+1)^2}\right]+\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{(t+1)^2}. \quad (2.7)$$

Next, substitute $|m|=r$ into equation (2.7) as

$$|g(m)|\geq\log(1-r)\left[\frac{-t^2+2tA_{\alpha\delta}-2A_{\alpha\delta}+1}{(t+1)^2}\right]+r\left[1+\left(\frac{-t^2+4tA_{\alpha\delta}+1}{(t+1)^2}\right)\right]$$

$$-\left(\frac{1}{1-r}\right)\left[\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{(t+1)^2}\right]+\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{(t+1)^2}. \quad (2.8)$$

Thus, multiply equation (2.8) with $\frac{1}{r}$ to obtain the equation for $\left|\frac{g(m)}{m}\right|$,

$$\left|\frac{g(m)}{m}\right| \le log(1-r)\left[\frac{-t^2+2tA_{\alpha\delta}-2A_{\alpha\delta}+1}{r(t+1)^2}\right]+\left[1+\left(\frac{-t^2+4tA_{\alpha\delta}+1}{(t+1)^2}\right)\right]$$
$$-\left(\frac{1}{1-r}\right)\left[\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{r(t+1)^2}\right]+\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{r(t+1)^2}$$

for $|m|=r<1$ as required.

This completes Theorem 2.1.

Based on Lemma 1.1, in Theorem 2.2, we are going to find a bound for $\frac{g(m)}{m}$ by using the $|a_n|=\frac{1}{n}\left(\frac{2A_{\alpha\delta}(n-1)+t+1}{(t+1)}\right)$ when n is odd.

Theorem 2.2

Let $g\in Z$, *since* $g\in S^*(\alpha,\delta,t)$ *and* $|a_n|=\frac{1}{n}\left(\frac{2A_{\alpha\delta}(n-1)+t+1}{(t+1)}\right)$ *where* $n=3,5,7,...,$ *then*

$$\log(1-r)\left[\frac{t+1-2A_{\alpha\delta}}{r(t+1)}\right]-\frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-r}\right)+2\le\left|\frac{g(m)}{m}\right|\le\log(1-r)\left[\frac{2A_{\alpha\delta}-t-1}{r(t+1)}\right]+\frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-r}\right)$$

for $|m|=r<1.$

Proof.

Given $g\in Z$, since $g\in\mathbb{N}$ satisfying Lemma 4.1.1 for $n=3,5,7,...$, we have

$$|g(m)|\le|m|+\sum_{n=2}^{\infty}|a_n||m|^n,$$

and

$$|g(m)|\le|m|+\sum_{n=2}^{\infty}\left|\frac{1}{n}\left(\frac{2A_{\alpha\delta}(n-1)+t+1}{(t+1)}\right)\right||m|^n \quad (2.9)$$

and the expansion of (2.9),

$$\left|g\left(m\right)\right| \le \left|m\right| + \sum_{n=2}^{\infty}\left|\frac{1}{n}\left(\frac{2nA_{\alpha\delta}}{\left(t+1\right)} - \frac{2A_{\alpha\delta}}{\left(t+1\right)} + \frac{t}{\left(t+1\right)} + \frac{1}{\left(t+1\right)}\right)\right|\left|m\right|^{n}$$

produce

$$\left|g\left(m\right)\right| \le \left|m\right| + \frac{2A_{\alpha\delta}}{\left(t+1\right)}\sum_{n=2}^{\infty}\left|m\right|^{n} - \frac{2A_{\alpha\delta}}{\left(t+1\right)}\sum_{n=2}^{\infty}\frac{\left|m\right|^{n}}{n} + \frac{t}{\left(t+1\right)}\sum_{n=2}^{\infty}\frac{\left|m\right|^{n}}{n} + \frac{1}{\left(t+1\right)}\sum_{n=2}^{\infty}\frac{\left|m\right|^{n}}{n} \qquad (2.10)$$

Then, simplification of (2.10) yields

$$\left|g\left(m\right)\right| \le \left|m\right| + \frac{2A_{\alpha\delta}}{\left(t+1\right)}\left(-1-\left|m\right| + \sum_{n=0}^{\infty}\left|m\right|^{n}\right) - \frac{2A_{\alpha\delta}}{\left(t+1\right)}\left(-\left|m\right| + \sum_{n=1}^{\infty}\frac{\left|m\right|^{n}}{n}\right)$$
$$+ \frac{t}{\left(t+1\right)}\left(-\left|m\right| + \sum_{n=1}^{\infty}\frac{\left|m\right|^{n}}{n}\right) + \frac{1}{\left(t+1\right)}\left(-\left|m\right| + \sum_{n=1}^{\infty}\frac{\left|m\right|^{n}}{n}\right)$$

Since $\log\left(1-\left|m\right|\right) = -\sum_{n=1}^{\infty}\frac{\left|m\right|^{n}}{n}$ and $\frac{1}{1-\left|m\right|} = \sum_{n=0}^{\infty}\left|m\right|^{n}$, thus

$$\left|g\left(m\right)\right| \le \left|m\right| + \frac{2A_{\alpha\delta}}{\left(t+1\right)}\left(-1-\left|m\right| + \frac{1}{1-\left|m\right|}\right) - \frac{2A_{\alpha\delta}}{\left(t+1\right)}\left(-\left|m\right| - \log\left(1-\left|m\right|\right)\right)$$
$$+ \frac{t}{\left(t+1\right)}\left(-\left|m\right| - \log\left(1-\left|m\right|\right)\right) + \frac{1}{\left(t+1\right)}\left(-\left|m\right| - \log\left(1-\left|m\right|\right)\right)$$

Which gives

$$\left|g\left(m\right)\right| \le \left|m\right| - \frac{2A_{\alpha\delta}}{\left(t+1\right)} + \frac{2A_{\alpha\delta}}{\left(t+1\right)}\left(\frac{1}{1-\left|m\right|}\right) + \frac{2A_{\alpha\delta}}{\left(t+1\right)}\left(\log\left(1-\left|m\right|\right)\right)$$
$$- \frac{t\left|m\right|}{\left(t+1\right)} - \frac{t}{\left(t+1\right)}\left(\log\left(1-\left|m\right|\right)\right) - \frac{\left|m\right|}{\left(t+1\right)} - \frac{1}{\left(t+1\right)}\left(\log\left(1-\left|m\right|\right)\right)$$

$$\left|g\left(m\right)\right| \le \log\left(1-\left|m\right|\right)\left[\frac{2A_{\alpha\delta}}{\left(t+1\right)} - \frac{t}{\left(t+1\right)} - \frac{1}{\left(t+1\right)}\right] + \left|m\right|\left[1 - \frac{t}{\left(t+1\right)} - \frac{1}{\left(t+1\right)}\right]$$
$$+ \left(\frac{1}{1-\left|m\right|}\right)\left[\frac{2A_{\alpha\delta}}{\left(t+1\right)}\right] - \frac{2A_{\alpha\delta}}{\left(t+1\right)} \qquad (2.11)$$

Next, substitute $\left|m\right| = r$ into equation (2.11) as

$$\left|g\left(m\right)\right| \le \log\left(1-r\right)\left[\frac{2A_{\alpha\delta} - t - 1}{\left(t+1\right)}\right] + \left(\frac{1}{1-r}\right)\left[\frac{2A_{\alpha\delta}}{\left(t+1\right)}\right] - \frac{2A_{\alpha\delta}}{\left(t+1\right)}. \qquad (2.12)$$

Thus, multiply equation (2.12) with $\frac{1}{r}$ and simplify to obtain the equation for $\left|\frac{g(m)}{m}\right|$,

$$\left|\frac{g(m)}{m}\right| \le \log(1-r)\left[\frac{2A_{\alpha\delta}-t-1}{r(t+1)}\right]+\frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-r}\right)$$

for $|m|=r<1$.

Hence, $g(m)$ absolutely converge in E, also $g(m)$ is analytic in E. Therefore, we h

$$|g(m)| \ge |m|-\sum_{n=2}^{\infty}|a_n||m|^n,$$

$$|g(m)| \ge |m|-\sum_{n=2}^{\infty}\left|\frac{1}{n}\left(\frac{2A_{\alpha\delta}(n-1)+t+1}{(t+1)}\right)\right||m|^n \tag{2.13}$$

and expansion of (2.13),

$$|g(m)| \ge |m|-\sum_{n=2}^{\infty}\left|\frac{1}{n}\left(\frac{2nA_{\alpha\delta}}{(t+1)}-\frac{2A_{\alpha\delta}}{(t+1)}+\frac{t}{(t+1)}+\frac{1}{(t+1)}\right)\right||m|^n$$

produces

$$|g(m)| \ge |m|-\frac{2A_{\alpha\delta}}{(t+1)}\sum_{n=2}^{\infty}|m|^n+\frac{2A_{\alpha\delta}}{(t+1)}\sum_{n=2}^{\infty}\frac{|m|^n}{n}-\frac{t}{(t+1)}\sum_{n=2}^{\infty}\frac{|m|^n}{n}-\frac{1}{(t+1)}\sum_{n=2}^{\infty}\frac{|m|^n}{n}. \tag{2.14}$$

Then, simplification of (2.14) yields

$$|g(m)| \ge |m|-\frac{2A_{\alpha\delta}}{(t+1)}\left(-1-|m|+\sum_{n=0}^{\infty}|m|^n\right)+\frac{2A_{\alpha\delta}}{(t+1)}\left(-|m|+\sum_{n=1}^{\infty}\frac{|m|^n}{n}\right)$$
$$-\frac{t}{(t+1)}\left(-|m|+\sum_{n=1}^{\infty}\frac{|m|^n}{n}\right)-\frac{1}{(t+1)}\left(-|m|+\sum_{n=1}^{\infty}\frac{|m|^n}{n}\right)$$

Since $\log(1-|m|)=-\sum_{n=1}^{\infty}\frac{|m|^n}{n}$ and $\frac{1}{1-|m|}=\sum_{n=0}^{\infty}|m|^n$, thus

$$|g(m)| \ge |m|-\frac{2A_{\alpha\delta}}{(t+1)}\left(-1-|m|+\frac{1}{1-|m|}\right)+\frac{2A_{\alpha\delta}}{(t+1)}\left(-|m|-\log(1-|m|)\right)$$
$$-\frac{t}{(t+1)}\left(-|m|-\log(1-|m|)\right)-\frac{1}{(t+1)}\left(-|m|-\log(1-|m|)\right)$$

Which gives

$$|g(m)| \geq |m| + \frac{2A_{\alpha\delta}}{(t+1)} - \frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-|m|}\right) - \frac{2A_{\alpha\delta}}{(t+1)}\left(\log(1-|m|)\right) + \frac{t|m|}{(t+1)} + \frac{t}{(t+1)}\left(\log(1-|m|)\right) + \frac{|m|}{(t+1)} + \frac{1}{(t+1)}\left(\log(1-|m|)\right)$$

$$|g(m)| \geq \log(1-|m|)\left[-\frac{2A_{\alpha\delta}}{(t+1)} + \frac{t}{(t+1)} + \frac{1}{(t+1)}\right] + |m|\left[1 + \frac{t}{(t+1)} + \frac{1}{(t+1)}\right] - \left(\frac{1}{1-|m|}\right)\left[\frac{2A_{\alpha\delta}}{(t+1)}\right] + \frac{2A_{\alpha\delta}}{(t+1)} \quad (2.15)$$

Next, substitute $|m| = r$ into equation (2.15) as

$$|g(m)| \geq \log(1-r)\left[\frac{t-2A_{\alpha\delta}+1}{(t+1)}\right] + 2r - \left(\frac{1}{1-r}\right)\left[\frac{2A_{\alpha\delta}}{(t+1)}\right] + \frac{2A_{\alpha\delta}}{(t+1)}. \quad (2.16)$$

Thus, multiply equation (2.16) with $\frac{1}{r}$ to obtain the equation for $\left|\frac{g(m)}{m}\right|$,

$$\left|\frac{g(m)}{m}\right| \geq \log(1-r)\left[\frac{t-2A_{\alpha\delta}+1}{r(t+1)}\right] + 2 - \left(\frac{1}{1-r}\right)\left[\frac{2A_{\alpha\delta}}{r(t+1)}\right] + \frac{2A_{\alpha\delta}}{r(t+1)}$$

$$\left|\frac{g(m)}{m}\right| \geq \log(1-r)\left[\frac{t+1-2A_{\alpha\delta}}{r(t+1)}\right] - \frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-r}\right) + 2$$

for $|m| = r < 1$ as required.

This completes Theorem 2.2.

Now, we shall prove our main result.

Main Results

Now, based on Theorem 2.1 and Theorem 2.2, we will apply these theorem to find the radius of starlikeness, $\Re_{St}$ for the class $S^*(\alpha,\delta,t)$. A similar approach of Nunokawa and Owa [4] and Yahya [6] will be applied to obtain the radius of starlikeness, $\Re_{St}$ for the class of $S^*(\alpha,\delta,t)$ in E.

Lemma 1.2 will be applied to obtain the result for Theorem 3.1 by using bound $|a_n| = \frac{1}{n}\left(\frac{1-t^2+2A_{\alpha\delta}\left[(-1+t)+n(t+1)\right]}{(t+1)^2}\right)$ where $n = 2,4,6,\ldots$.

Theorem 3.1

Let $g \in S^*(\alpha,\delta,t)$ and g is univalent and starlike in $|m| < \Re_{St}$, then radius of starlikeness, $\Re_{St}$ is given

$$log\left(1-r-\sqrt{2}r\right)\left[\frac{t^2-2tA_{\alpha\delta}+2A_{\alpha\delta}-1}{\left(r+\sqrt{2}r\right)(t+1)^2}\right]+\left[1+\left(\frac{t^2-4tA_{\alpha\delta}-1}{(t+1)^2}\right)\right]$$
$$+\left(\frac{1}{1-r-\sqrt{2}r}\right)\left(\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{\left(r+\sqrt{2}r\right)(t+1)^2}\right)-\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{\left(r+\sqrt{2}r\right)(t+1)^2}-e.$$

Proof.

By means of Theorem 2.1, we have $\left|\frac{g(m)}{m}\right| > 0$ in $|m| < R_1 = 0.7475842618$ and therefore, $log\left(\frac{g(m)}{m}\right)$ is harmonic in $|m| < R$. From Lemma 1.2 and based on the harmonic function theory, we defined

$$log\frac{g(m)}{m} = \frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\frac{\zeta+m}{\zeta-m}d\phi$$

where $\zeta = \rho e^{i\phi},(0 \le \phi \le 2\pi), m = re^{i\theta},(0 \le \theta \le 2\pi)$, and $0 \le r < \rho \le R_1 = 0.7475842618$.

Logarithmic differentiation approach will be applied and yields

$$\frac{d}{dm}\left(log\frac{g(m)}{m}\right) = \frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\frac{d}{dm}\left(\frac{\zeta+m}{\zeta-m}\right)d\phi$$

and

$$\frac{d}{g(m)}\left[\frac{g'(m)}{m}-\frac{g(m)}{m^2}\right] = \frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\left(\frac{(\zeta-m)+(\zeta+m)}{(\zeta-m)^2}\right)d\phi. \quad (3.1)$$

Then we simplify of (3.1) will results

$$\frac{1}{m}\left[\frac{mg'(m)}{g(m)}-1\right] = \frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\left(\frac{2\zeta}{(\zeta-m)^2}\right)d\phi$$

and

$$\frac{mg'(m)}{g(m)} = 1 + \frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\left(\frac{2m\zeta}{(\zeta-m)^2}\right)d\phi. \tag{3.2}$$

From (3.2), we have

$$\frac{mg'(m)}{g(m)} = 1 + \frac{1}{2\pi}\int_0^{2\pi}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\left(\left|\frac{2m\zeta}{(\zeta-m)^2}\right|\right)d\phi$$

and yields

$$\frac{mg'(m)}{g(m)} = 1 + \frac{1}{2\pi}\int_0^{2\pi}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\left(\frac{2\left|re^{i\theta}\right|\left|\rho e^{i\phi}\right|}{\left|\rho e^{i\phi}-re^{i\theta}\right|^2}\right)d\phi.$$

Let $\left|\rho e^{i\phi}\right| = \rho$ and $\left|re^{i\theta}\right| = r$, we have

$$\frac{mg'(m)}{g(m)} = 1 + \frac{1}{2\pi}\int_0^{2\pi}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\frac{2r\rho}{\left|(\rho\cos\phi - r\cos\theta) + i(\rho\sin\phi - r\sin\theta)^2\right|}d\phi$$

which gives

$$\frac{mg'(m)}{g(m)} = 1 + \frac{1}{2\pi}\int_0^{2\pi}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\frac{2r\rho}{(\rho\cos\phi - r\cos\theta)^2 + (\rho\sin\phi - r\sin\theta)^2}d\phi$$

So that,

$$\frac{mg'(m)}{g(m)} = 1 + \frac{1}{2\pi}\int_0^{2\pi}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\frac{2r\rho}{\rho^2 - 2\rho r\cos(\rho-\theta) + r^2}d\phi$$

From Theorem 2.1, we have

$$log(1-r)\left[\frac{-t^2 + 2tA_{\alpha\delta} - 2A_{\alpha\delta} + 1}{r(t+1)^2}\right] + \left[1 + \left(\frac{-t^2 + 4tA_{\alpha\delta} + 1}{(t+1)^2}\right)\right] - \left(\frac{1}{1-r}\right)\left(\frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{r(t+1)^2}\right)$$
$$+ \frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{r(t+1)^2} < log(1-r)\left[\frac{t^2 - 2tA_{\alpha\delta} + 2A_{\alpha\delta} - 1}{r(t+1)^2}\right] + \left[1 + \left(\frac{t^2 - 4tA_{\alpha\delta} - 1}{(t+1)^2}\right)\right]$$
$$+ \left(\frac{1}{1-r}\right)\left(\frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{r(t+1)^2}\right) - \frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{r(t+1)^2}$$

for $|m| = r < 1$, we derive

$$\mathrm{Re}\left\{\frac{mg'(m)}{g(m)}\right\} \geq 1 - \frac{1}{2\pi}\int_0^{2\pi} max_{|\zeta|=\rho}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\frac{2\rho r}{\rho^2 - 2\rho\, cos(\phi-\theta) + r^2}\, d\phi \tag{3.3}$$

and from (3.3) and integrating with respect to ϕ yield

$$\mathrm{Re}\left\{\frac{mg'(m)}{g(m)}\right\} \geq 1 - \frac{2\rho r}{\rho^2 - r^2} log\left(l\,og(1-\rho)\left[\frac{t^2 - 2tA_{\alpha\delta} + 2A_{\alpha\delta} - 1}{\rho(t+1)^2}\right]\right.$$
$$\left. + \left[1 + \left(\frac{t^2 - 4tA_{\alpha\delta} - 1}{(t+1)^2}\right)\right] + \left(\frac{1}{1-\rho}\right)\left(\frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{\rho(t+1)^2}\right) - \frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{\rho(t+1)^2}\right) \tag{3.4}$$

where $0 \leq r < \rho \leq R_1 = 0.7475842618$.

Based on $\mathrm{Re}\left\{\frac{mg'(m)}{g(m)}\right\} \geq 0$ in $|m| < \Re_{St}$ and from (3.4), we see that

$$\frac{2\rho r}{\rho^2 - r^2} log\left(l\,og(1-\rho)\left[\frac{t^2 - 2tA_{\alpha\delta} + 2A_{\alpha\delta} - 1}{\rho(t+1)^2}\right] + \left[1 + \left(\frac{t^2 - 4tA_{\alpha\delta} - 1}{(t+1)^2}\right)\right]\right.$$
$$\left. + \left(\frac{1}{1-\rho}\right)\left(\frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{\rho(t+1)^2}\right) - \frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{\rho(t+1)^2}\right) = 1$$

and

$$log\left(l\,og(1-\rho)\left[\frac{t^2 - 2tA_{\alpha\delta} + 2A_{\alpha\delta} - 1}{\rho(t+1)^2}\right] + \left[1 + \left(\frac{t^2 - 4tA_{\alpha\delta} - 1}{(t+1)^2}\right)\right]\right.$$
$$\left. + \left(\frac{1}{1-\rho}\right)\left(\frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{\rho(t+1)^2}\right) - \frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{\rho(t+1)^2}\right) = \frac{\rho^2 - r^2}{2\rho r}. \tag{3.5}$$

Putting $\rho = \left(1+\sqrt{2}\right)r$ in (3.5), we have

$$log\left(l\,og\left(1-\left(1+\sqrt{2}\right)r\right)\left[\frac{t^2 - 2tA_{\alpha\delta} + 2A_{\alpha\delta} - 1}{\left(\left(1+\sqrt{2}\right)r\right)(t+1)^2}\right] + \left[1 + \left(\frac{t^2 - 4tA_{\alpha\delta} - 1}{(t+1)^2}\right)\right] + \left(\frac{1}{1-\left(1+\sqrt{2}\right)r}\right)\right.$$
$$\left.\left(\frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{\left(\left(1+\sqrt{2}\right)r\right)(t+1)^2}\right) - \frac{2tA_{\alpha\delta} + 2A_{\alpha\delta}}{\left(\left(1+\sqrt{2}\right)r\right)(t+1)^2}\right) = \frac{\left(\left(1+\sqrt{2}\right)r\right)^2 - r^2}{2\left(\left(1+\sqrt{2}\right)r\right)r}$$

and

$$log\left(l\,og\left(1-r-\sqrt{2}r\right)\left[\frac{t^2-2tA_{\alpha\delta}+2A_{\alpha\delta}-1}{\left(r+\sqrt{2}r\right)\left(t+1\right)^2}\right]+\left[1+\left(\frac{t^2-4tA_{\alpha\delta}-1}{\left(t+1\right)^2}\right)\right]\right.$$
$$\left.+\left(\frac{1}{1-r-\sqrt{2}r}\right)\left(\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{\left(r+\sqrt{2}r\right)\left(t+1\right)^2}\right)-\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{\left(r+\sqrt{2}r\right)\left(t+1\right)^2}\right)=1$$

which gives

$$l\,og\left(1-r-\sqrt{2}r\right)\left[\frac{t^2-2tA_{\alpha\delta}+2A_{\alpha\delta}-1}{\left(r+\sqrt{2}r\right)\left(t+1\right)^2}\right]+\left[1+\left(\frac{t^2-4tA_{\alpha\delta}-1}{\left(t+1\right)^2}\right)\right]$$
$$+\left(\frac{1}{1-r-\sqrt{2}r}\right)\left(\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{\left(r+\sqrt{2}r\right)\left(t+1\right)^2}\right)-\frac{2tA_{\alpha\delta}+2A_{\alpha\delta}}{\left(r+\sqrt{2}r\right)\left(t+1\right)^2}=e$$

$|m|<\Re_{St}$.

This completes Theorem 3.1.

Next, Lemma 1.2 will be applied to obatin the result for Theorem 3.2 by using bound $|a_n|=\frac{1}{n}\left(\frac{2A_{\alpha\delta}\left(n-1\right)+t+1}{\left(t+1\right)}\right)$ where $n=3,5,7,....$

Theorem 3.2

Let $g\in S^*\left(\alpha,\delta,t\right)$, and g is univalent and starlike in $|m|<\Re_{St}$, then radius of starlikeness, $\Re_{St}$ is given

$$\frac{2A_{\alpha\delta}}{\left(t+1\right)}\left(\frac{1}{1-r-\sqrt{2}r}\right)+\left(\frac{2A_{\alpha\delta}-t-1}{\left(r+\sqrt{2}r\right)\left(t+1\right)}\right)l\,og\left(1-r-\sqrt{2}r\right)-e$$

Proof.

Through Theorem 2.2, we have $\left|\frac{g\left(m\right)}{m}\right|>0$ in $|m|<R_1=0.7475842618$ and therefore, $l\,og\left(\frac{g\left(m\right)}{m}\right)$ is harmonic in $|m|<R$. From Lemma 1.2 and based on the harmonic function theory, we defined

$$log\frac{g(m)}{m}=\frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\frac{\zeta+m}{\zeta-m}d\phi$$

where $\zeta=\rho e^{i\phi},(0\le\phi\le 2\pi), m=re^{i\theta},(0\le\theta\le 2\pi)$, and $0\le r<\rho\le R_1=0.7475842618$.

Logarithmic differentiation approach will be applied and yields

$$\frac{d}{dm}\left(log\frac{g(m)}{m}\right)=\frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\frac{d}{dm}\left(\frac{\zeta+m}{\zeta-m}\right)d\phi$$

and

$$\frac{d}{g(m)}\left[\frac{g'(m)}{m}-\frac{g(m)}{m^2}\right]=\frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\left(\frac{(\zeta-m)+(\zeta+m)}{(\zeta-m)^2}\right)d\phi. \tag{3.6}$$

Then we simplify of (3.6) will results

$$\frac{1}{m}\left[\frac{mg'(m)}{g(m)}-1\right]=\frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\left(\frac{2\zeta}{(\zeta-m)^2}\right)d\phi$$

and

$$\frac{mg'(m)}{g(m)}=1+\frac{1}{2\pi}\int_0^{2\pi}\left(log\left|\frac{g(\zeta)}{\zeta}\right|\right)\left(\frac{2m\zeta}{(\zeta-m)^2}\right)d\phi. \tag{3.7}$$

From (3.7), we have

$$\frac{mg'(m)}{g(m)}=1+\frac{1}{2\pi}\int_0^{2\pi}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\left(\left|\frac{2m\zeta}{(\zeta-m)^2}\right|\right)d\phi$$

and yields

$$\frac{mg'(m)}{g(m)}=1+\frac{1}{2\pi}\int_0^{2\pi}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\left(\frac{2\left|re^{i\theta}\right|\left|\rho e^{i\phi}\right|}{\left|\rho e^{i\phi}-re^{i\theta}\right|^2}\right)d\phi.$$

Let $\left|\rho e^{i\phi}\right|=\rho$ and $\left|re^{i\theta}\right|=r$, we have

$$\frac{mg'(m)}{g(m)}=1+\frac{1}{2\pi}\int_0^{2\pi}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\frac{2r\rho}{\left|(\rho\cos\phi-r\cos\theta)+i(\rho\sin\phi-r\sin\theta)^2\right|}d\phi$$

which gives

$$\frac{mg'(m)}{g(m)} = 1 + \frac{1}{2\pi}\int_0^{2\pi}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\frac{2r\rho}{(\rho\cos\phi - r\cos\theta)^2 + (\rho\sin\phi - r\sin\theta)^2}d\phi .$$

From Theorem 2.2, we have

$$log(1-r)\left[\frac{2A_{\alpha\delta} - t - 1}{r(t+1)}\right] + \frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-r}\right) + 2 < log(1-r)\left(\frac{2A_{\alpha\delta} - t - 1}{r(t+1)}\right) + \frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-r}\right)$$

for $|m| = r < 1$, we derive

$$\mathrm{Re}\left\{\frac{mg'(m)}{g(m)}\right\} \geq 1 - \frac{1}{2\pi}\int_0^{2\pi} max_{|\zeta|=\rho}\left(\left|log\left|\frac{g(\zeta)}{\zeta}\right|\right|\right)\frac{2\rho r}{\rho^2 - 2\rho\cos(\phi-\theta) + r^2}d\phi \tag{3.8}$$

and from (3.8) and integrating with respect to ϕ yield

$$\mathrm{Re}\left\{\frac{mg'(m)}{g(m)}\right\} \geq 1 - \frac{2\rho r}{\rho^2 - r^2}log\left(log(1-\rho)\left(\frac{2A_{\alpha\delta} - t - 1}{\rho(t+1)}\right) + \frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-\rho}\right)\right) \tag{3.9}$$

where $0 \leq r < \rho \leq R_1 = 0.7475842618$.

Based on $\mathrm{Re}\left\{\frac{mg'(m)}{g(m)}\right\} \geq 0$ in $|m| < \Re_{St}$ and from (3.9), we see that

$$log\left(log(1-\rho)\left(\frac{2A_{\alpha\delta} - t - 1}{\rho(t+1)}\right) + \frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-\rho}\right)\right) = \frac{\rho^2 - r^2}{2\rho r}. \tag{3.10}$$

Putting $\rho = \left(1+\sqrt{2}\right)r$ in (3.10), we have

$$log\left(log\left(1-\left(1+\sqrt{2}\right)r\right)\left(\frac{2A_{\alpha\delta} - t - 1}{\left(\left(1+\sqrt{2}\right)r\right)(t+1)}\right) + \frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-\left(1+\sqrt{2}\right)r}\right)\right) = \frac{\left(\left(1+\sqrt{2}\right)r\right)^2 - r^2}{2\left(\left(1+\sqrt{2}\right)r\right)r}$$

and

$$log\left(log\left(1-r-\sqrt{2}r\right)\left(\frac{2A_{\alpha\delta} - t - 1}{\left(r+\sqrt{2}r\right)(t+1)}\right) + \frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-r-\sqrt{2}r}\right)\right) = 1$$

which gives

$$\frac{2A_{\alpha\delta}}{(t+1)}\left(\frac{1}{1-r-\sqrt{2}r}\right)+\left(\frac{2A_{\alpha\delta}-t-1}{\left(r+\sqrt{2}r\right)(t+1)}\right)log\left(1-r-\sqrt{2}r\right)=e$$

in $|m|<\Re_{St}$.

This completes Theorem 3.2.

Conclusion

In conclusion, there is a purpose of this paper, which is to determine the radius of starllikenss, $\Re_{St}$ using the Poisson formula. We believe that we have achieved the objective that we highlighted.

References

[1] Goodman, A. W. (1983). Univalent functions. Polygonal Publishing House, Washington, New Jersey, Vol. I & II.

[2] Duren, P. L. (1983). Univalent Function.*Springer Verlag New York Inc*

[3] Rathi, S. (2015) Properties of titled univalent analytics function of order δ, Unpublished Master Thesis, Universiti Teknologi MARA.

[4] Nunokawa, M., & Owa, S. (2000). On some inverse properties for univalent functions. Kyoto University Koukyuroku, 1164, 73-76.

[5] Ikeda, & Saigo, M. (2000). Various inverse problems for univalent functions. RIMS Kokyu Record, 1164, 22-30.

[6] Yahya, A., & Soh, S. C. (2017). Radius of starlikeness of certain class of close-to-convex function. International Journal of Pure and Applied Mathematical, 10, 341-348. http://dx.doi.org/10.12732/ijpam.v112i2.11

NUMERICAL SOLUTION OF NANOFLUID FLOW OVER A PERMEABLE STRETCHING AND SHRINKING SURFACE USING BVP4C SOLVER

(exactly 12 pt)

Nurqasrina Ahmad Azian[1,*], Siti Hidayah Muhad Saleh[2]

[1]*College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Kampus Seremban, Negeri Sembilan*

[*]qasrina.azian02@gmail.com

Abstract

This study investigates nanofluid flow over a permeable stretching/shrinking surface using BVP4C. This study utilises copper and silver nanoparticles to examine how well the two nanoparticles transfer heat. The aim of this study is to apply the similarity transformation to convert the partial differential equation (PDE) to an ordinary differential equation (ODE). The objective is to solve nanofluid flow over a permeable stretching/shrinking surface by using BVP4C. BVP4C in MATLAB software is then used to solve this equation numerically. The study then compares the results between the previous method which is the shooting method , and BVP4C, thus identifying the impact of copper and silver nanoparticles on the stretching/shrinking surface that influence the efficiency of heat transfer by analysing and measuring the nanoparticles in velocity profile, temperature profile, skin friction coefficient and local Nusselt number. The comparison demonstrates strong agreement between the computed skin friction coefficient and local Nusselt number obtained using BVP4C. The obtained result indicates that in the stretching case, the velocity profile, temperature profile, and boundary layer thickness decrease, but in the shrinking case, the velocity profile, temperature profile, and boundary layer thickness increase. In addition, the skin friction coefficient of silver is higher than that of copper, which contributes to a decrease in the velocity boundary layer. Meanwhile, the local Nusselt number of silver nanoparticles are higher than that of copper, indicating a higher rate of heat transfer for silver.

Keywords*:* Nanofluid, stretching/shrinking surfaces, permeable, BVP4C

1. Introduction

The concept of nanofluid emerged with the intention of improving the fluid's thermal conductivity, which encouraged advancements in a range of engineering sectors. The discovery of heat transmission fluids like air, ethylene glycol, water, and oil have extremely poor thermal conductivity compared to solids. The idea of renewal led to developing the electrical conductivity of heterogeneous solid particles and allowed the study of the conductivity of a mixture of liquid and solid particles to increase thermal conductivity [1]. The purpose of nanofluid remains the same intent of emphasising the improvement of thermal characteristics to the new technology specifically, nanotechnology in order to attain the greatest possible thermal conductivity. A permeable surface is a surface with holes or boundaries that let fluid

pass through or be absorbed. The surface has the ability for the fluid to be suction or injected into the boundary layer of the surface. Suction is a method used to manage flow and modify turbulence induced by flow instability downstream of the boundary layer separation point [2]. The fluid flow over a permeable surface is commonly used in engineering applications, including air foil design, filtration systems, and combustion chambers.

Next, the structure of the boundary layer, which is defined as a light layer of gas or liquid that touches an object's surface, is affected by the existence of suspended nanoparticles in the general fluid. Permeable surface and boundary layer have a significant influence on heat transfer [3]. Heat transmission rates are enhanced when fluid is removed from the boundary layer because convective processes are more efficient [4]. While an injection of the surface will reduce the transmission of heat since the heat of the surface has dropped due to the additional fluid. Therefore, heat transmission is a basic thermodynamic and engineering concept that describes the motion of the heat energy from one system to another as due to temperature differences. Conduction, convection, and radiation are the three primary processes in heat transmission with different circumstances. In addition, this research uses copper and silver as a nanoparticle to study the impact on the transfer of heat and fluid flow. Several studies have established that nanofluids provide an effective way of conveying thermal conductivity and heat.

The impact of surface roughness on mixed convective nanofluid flow across an exponentially extending permeable surface using the quasilinearization technique has been conducted by [5]. The research findings conclude that nanoparticles in the base fluid have been significantly affected by its flow since it increases friction around walls and reduces heat transfer from hot to cold fluid. The study made by [6] for nanofluid flow past a nonlinearly stretching surface with shear flow and zero nanoparticle flux by the shooting method, indicated that the injection surface will affect temperature drops and wall shear stress and decrease in heat transfer speed. The velocity of fluid reduced when the nonlinear stretching parameter was amplified. The Prandtl number helps in cooling in a variety of systems when the heat of boundary layer width decreases. The numerical solution of nanofluid flow over permeable stretching and shrinking surfaces has been well researched. From the previous study, nanofluids exhibit superior thermal conductivity compared to base fluids, enhancing the rate of heat transfer between solid particles and the base fluid due to increased surface area [1].

The study of [7] stated that the heat transfer rate increases with heat radiation. Boundary layer separation was delayed by suction parameter growth while the increase in volume amount of hybrid nanofluids (copper and alumina) provides superior heat transport compared to nanofluids and water. While research by [8] studied about examining how TiO2-Ag/water affects hybrid nanofluid MHD flow while taking radiation and slip conditions into account across exponentially expanding and contracting sheets by using the BVP4C method. The research concludes that the presence of the nanoparticles raises the temperature. When the velocity slip parameter was raised, the skin-friction coefficient reduced in the $\lambda < 0$ shrinking surface region but contradiction was seen on the $\lambda > 0$ stretching surface, while the Nusselt number had the opposite impact. On the other hand, the results showed that increases in skin friction and Nusselt number were influenced by the volume of nanoparticles and the heat

transfer rate when a general nanofluid was transformed into a hybrid nanofluid by raising its viscosity. Higher skin friction coefficient and velocity gradient are the results of increased instability parameters [9].

The study of a three-dimensional unsteady radiative hybrid nanofluid flow over a permeable surface through a porous space by the BVP4C method concludes that the increase in heat transfer rate has the potential to decrease shrinking capacity, while excessive surface shrinking might raise the Skin Friction Coefficient [2]. Hence, the study of the rate of heat transfer rises as a result of an unstable hybrid nanofluid flow across a radially permeable shrinking/stretching surface, depending on the mass suction parameter and the volume percentage of nanoparticles [10]. Lastly, the research on the effect of radiation from heat and suction effect the unstable ternary mixed nanofluid flow across a biaxial shrinking sheet indicates that shows that raising the suction parameter in ternary innovative hybrid nanofluids generates improvement in local skin friction and a delay in boundary layer separations. Boundary layer separation was not significantly influenced by thermal radiation. Higher velocities and a lower system temperature are the results of increased suction. While thermal radiation at high intensities significantly increases the ternary hybrid nanofluid system's
internal temperature [11].

Thus, the focus of this study is to find a different approach to solving the nanofluid flow problem to the stretching and shrinking surfaces while investigating it with MATLAB's BVP4C.

2. Methodology

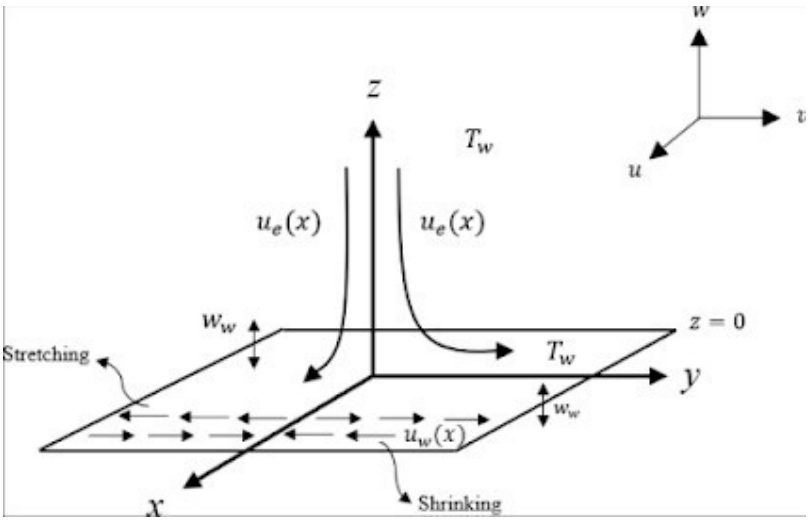


Figure 1: Physical model of stretching/shrinking surface

The figure above illustrates the model of the study that uses several variables and notations that have their own value and effect. The stretching and shrinking sheet denoted as $u_w(x)$ while

$u_e(x)$ is the outward flow velocity of the nanofluid with the constant c, where $c > 0$ is for the stretching surface while $c < 0$ for the shrinking surface [13]. Based on Figure 1, moving outward-pointing arrows show the surface is stretching, while inward-pointing arrows show the shrinking surface. On the other hand, the velocity of mass flux is denoted as w_w . The injection of mass flux occurs when $w_w > 0$ while $w_w < 0$ for suction. Next, the notation of T_w is the temperature of fluid in the boundary layer since this researcher used different fluids which are water-based either copper (Cu) or silver (Ag). Hence,

The continuity equation is governed by,

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{1}$$

The momentum equation is,

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = u_e\frac{\partial u_e}{\partial x} + v_{nf}\frac{\partial^2 u}{\partial z^2} \tag{2}$$

Energy equation governed by,

$$u\frac{\partial T}{\partial x} + u\frac{\partial T}{\partial x} + u\frac{\partial T}{\partial x} = \alpha_{nf}\frac{\partial^2 T}{\partial z^2} \tag{3}$$

and the boundary conditions are

$$\begin{aligned} &u = u_w(x) = ax, \quad v = 0, \quad w = w_w, \quad T = T_w \quad at\ z = 0 \\ &u = u_e(x) = cx, \quad T = T_\infty \quad at\ z = \infty \end{aligned} \tag{4}$$

where variable u and v are the components of x and y velocity respectively, $u_w(x) = ax$ is referred to shrinking/stretching velocity and $u_e(x) = cx$ denoted as the velocity of nanofluid external flow. This parameter will stretch the surface if the value of $c > 0$ while shrinking is $c < 0$. Then, w_w denoted as the velocity of mass flux , where it is injection when the condition $w_w > 0$ and suction when $w_w < 0$. The use of the parameter T represents temperature where T_w is the surface temperature and T_∞ is the temperature of the fluid far away from the surface. Therefore,

The similarity equation variables given by

$$u = cf'(\eta)\,, \qquad v = c(m-1)yf'(\eta), \qquad w = -\sqrt{cv}\,mf(\eta)$$
$$\theta(\eta) = \frac{T - T_\infty}{T_w - T_\infty}, \qquad \eta = \sqrt{\frac{c}{v}}\,z \tag{5}$$

In this case, $m = 1$ represents a single direction of shrinking sheet and $m = 2$ an asymmetric shrinkage. Hence, by using a similarity transformation, the differentiated variable will be substituted into (1), (2), and (3) to obtain the ODE as below

$$k_1 f''' + mf\,f'' + 1 - f'^2 = 0 \tag{6}$$

and

$$k_2\,\theta'' + Prmf\,\theta' = 0 \tag{7}$$

with boundary condition as follows

$$f(0) = s,\ f'(0) = \lambda,\ \theta(0) = 1,\ f'(\eta) \to 1,\ \theta(\eta) \to 0 \quad as\ \eta \to \infty \tag{8}$$

where k_1, k_2, the Prandtl number (Pr), shrinking and stretching parameter (λ) and suction parameter (s) are defined as:

$$k_1 = \frac{1}{(1-\phi)^{2.5}\left(\frac{1-\phi+\phi\rho_s}{\rho_f}\right)}, \quad k_2 = \frac{k_{nf}/k_f}{(1-\phi)+\left(\frac{\phi(\rho C_\rho)_s}{(\rho C_\rho)_f}\right)}, \tag{9}$$

$$Pr = \frac{v_f}{\alpha_f}, \quad s = -\frac{w_w}{(cv)^{\frac{1}{2}}}, \quad \lambda = \frac{\alpha}{c}, \tag{10}$$

where the suction parameter is stretching when $\lambda > 0$ while shrinking $\lambda < 0$.

This study only focuses on two physical quantities, which are the skin friction coefficient, C_f and the local Nusselt number, Nu_x which both of which can be written as equation below:

$$Re_x^{\frac{1}{2}} C_f = \frac{1}{(1-\phi)^{2.5}} f''(0), \quad Re_x^{-\frac{1}{2}} Nu_x = -\frac{k_{nf}}{k_f}\theta'(0) \tag{11}$$

where the dimensionless parameter of Reynolds number as $Re_x = \frac{u_e(x)x}{v}$. The effective viscosity and thermal diffusivity of nanofluid. Hence, the variable was given as below.

$$\mu_{nf} = \frac{\mu_f}{(1-\phi)^{2.5}}, \quad \alpha_{nf} = \frac{k_{nf}}{\left(\rho C_\rho\right)}, \quad \frac{k_{nf}}{k_f} = \frac{k_s + 2k_f - 2\phi\left(k_f - k_s\right)}{k_s + 2k_f - \phi\left(k_f - k_s\right)} \tag{12}$$

and

$$\left(\rho C_\rho\right)_{nf} = (1-\phi)\left(\rho C_\rho\right)_f + \phi\left(\rho C_\rho\right)_{s,} \quad \rho_{nf} = (1-\phi)_{\rho f} + \phi\rho_s. \tag{13}$$

3. Result and discussion

The results are compared with two previous findings by [13], who applied the shooting method with Runge-Kutta Fehlberg in Maple and [12], who solved the problem numerically using another method. This comparison aims to confirm that the results generated by BVP4C are valid. Table 1 represents the comparison results for skin friction coefficient, $f''(0)$, when $s = 0,\ \phi = 0$ with various of λ for Cu nanoparticle with $m = 1, 2$. While Table 2 presents the comparison of the value local Nusselt number for Cu nanoparticle with $m = 2, s = 0.5$ and various of λ.

Table 1: The comparison of value skin friction coefficient $f''(0)$ for Cu nanoparticle with $m = 1, 2$ (the sheet shrink at one direction and asymmetrically) $s = 0, \phi = 0$ and various of λ.

λ	[13]		[12]		Present study	
	$m = 1$	$m = 2$	$m = 1$	$m = 2$	$m = 1$	$m = 2$
1	0	0	0	0	0	0
0.5	-	-	0.1733	0.78032	0.7133	0.78032
0	1.232588	1.311938	1.232588	1.311938	1.232588	1.311938
-0.5	-	-	1.49567	1.49001	1.49567	1.49001
-0.75	-	-	1.49567	1.49001	1.49567	1.49001

Table 2: The comparison of local Nusselt number for Cu nanoparticle with $m = 2, s = 0.5$ and various of λ.

λ	[12]		Present study	
	$\phi = 0$	$\phi = 0.1$	$\phi = 0$	$\phi = 0.1$
1	8.2292	6.3563	8.2292	6.3563
0.5	7.7302	5.9489	7.7302	5.9489
0	7.1440	5.4757	7.1440	5.4757
-0.5	6.4033	4.8927	6.4033	4.8927
-0.75	5.9206	4.5291	5.9206	4.5291

Based on the table comparison of Table 1 and Table 2 the value of skin friction and Nusselt number obtained as previous study. Thus, this indicates that the present study is consistent with the previous research even though the method used is different. The previous study obtained the result by using shooting method whereas the present study used BVP4C method. Hence, the research satisfies the agreement.

3.1 Effect of shrinking/stretching parameter, λ on velocity profile and temperature profile, skin friction coefficient and local Nusselt number.

According to Figure 3.1 and Figure 3.2, it shows the different values of λ on temperature and velocity profile respectively. Figure 3.1 represents the velocity profile graft while Figure 3.2 representes the temperature profile with different values of stretching/shrinking value of λ. The negative value of $\lambda = -1.5, -1.6$ shows there exists a second solution which is considered as shrinking cases while there is no dual solution in the stretching case. Hence, the velocity profile illustrates that the stretching case increases steadily as it approaches 1. This reduces the thickness of the boundary layer. While Figure 3.2 demonstrates that the temperature profile declines with different values of lambda. However, stretching generates a reduction in the boundary layer as heat transfer from the surface to the fluid increases.

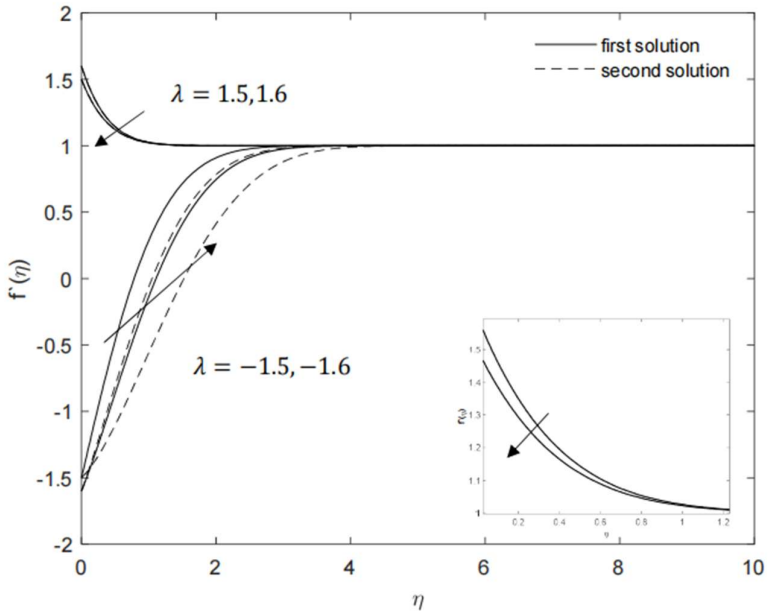


Figure 3.1: The velocity profile for different types of stretching/shrinking parameter $\lambda = 1.5,1.6,-1.5,-1.6$ with $m = 1, s = 0.5$ (permeable).

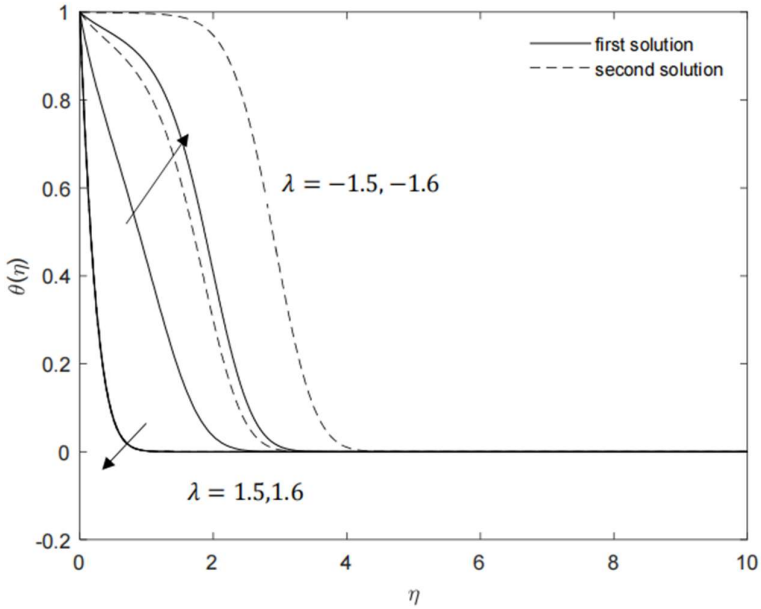


Figure 3.2: The temperature profile for different types of stretching/shrinking parameter $\lambda = 1.5,1.6,-1.5,-1.6$ with $m = 1, s = 0.5$.

Figure 3.3 and Figure 3.4 below illustrate the graph of skin friction coefficient and local Nusselt number that shows the dual solution exists based on the profile graph above. Figure 3.3 shows that for $\lambda > 1$, as the rise of the suction parameter s contributes to a higher value of the skin friction coefficient, thus it is indicated that the suction delays the boundary layer separation.

While Figure 3.4, the increase of local Nusselt number as suction parameter also increases where λ is decreases. Thus, it is illustrated that the suction delayed separation of boundary layer. In addition, both Figure 3.3 and Figure 3.4 demonstrate that the second solution exists in $s = 0.5, 0.6, 0.7$. A difference in the stretching/shrinking parameter, λ indicates the existence of the second solution where it only exists at the negative value for all values of s which shows it exist in the shrinking cases but not in stretching cases.

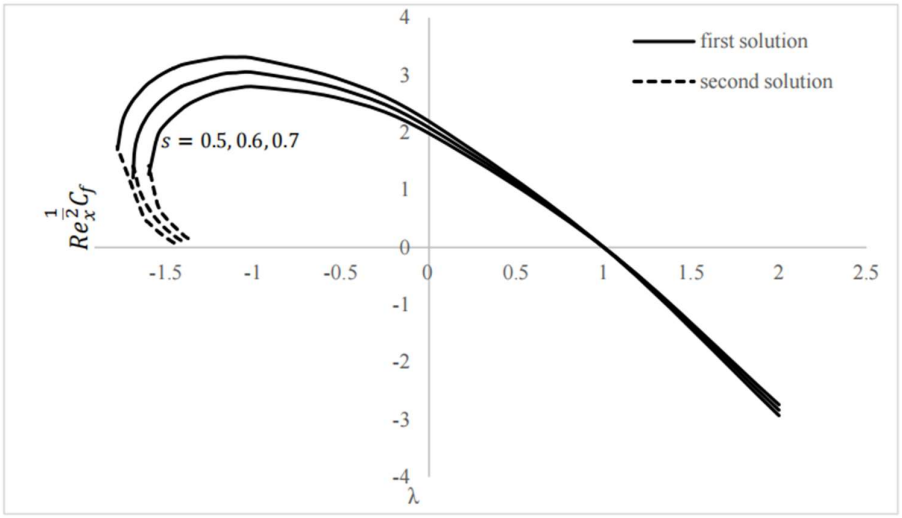


Figure 3.3: The Skin friction coefficient, $Re_x^{\frac{1}{2}} C_f$ versus λ with different type suction parameter, $s = 0.5, 0.6, 0.7$ when $m = 1, \phi = 0.1$.

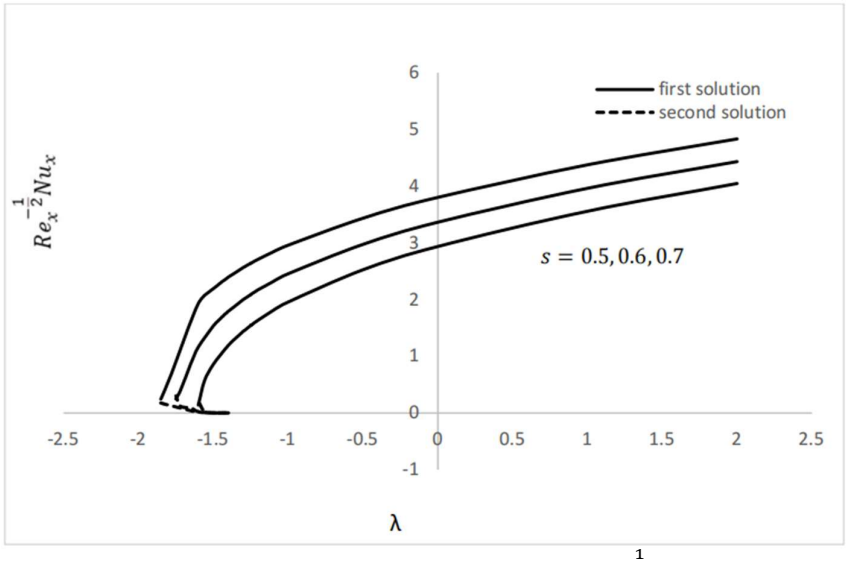


Figure 3.4: Variation of local Nusselt number, $Re_x^{-\frac{1}{2}} Nu_x$ versus λ for Ag nanoparticle when $m = 1, \phi = 0.1$ and various value of s

3.2 Effect of different types of nanoparticles (copper and silver) on skin friction coefficient and local Nusselt number.

The skin friction coefficient and local Nusselt number versus λ with different types of nanoparticles (Cu, Ag) when $m = 1, \phi = 0.1, s = 0.5$ were shown in Figure 3.5 and Figure 3.6 below. From Figures 3.5 and 3.6 there exist regions with unique solutions for $\lambda > -1$, dual solutions for $\lambda_c < \lambda < -1$, and no solutions for $\lambda < \lambda_c$, where λ_c represents the critical value of λ. It is demonstrated that as λ drops, the skin friction coefficient increases and the local Nusselt number decreases. Therefore, Figure 3.6 shows the local Nusselt number for copper and silver decreases as lambda decreases. The transfer of heat from the fluid's surface is measured by the Nusselt number where increased Nusselt number, enhanced heat transportation, and reduced thermal boundary layer thickness.

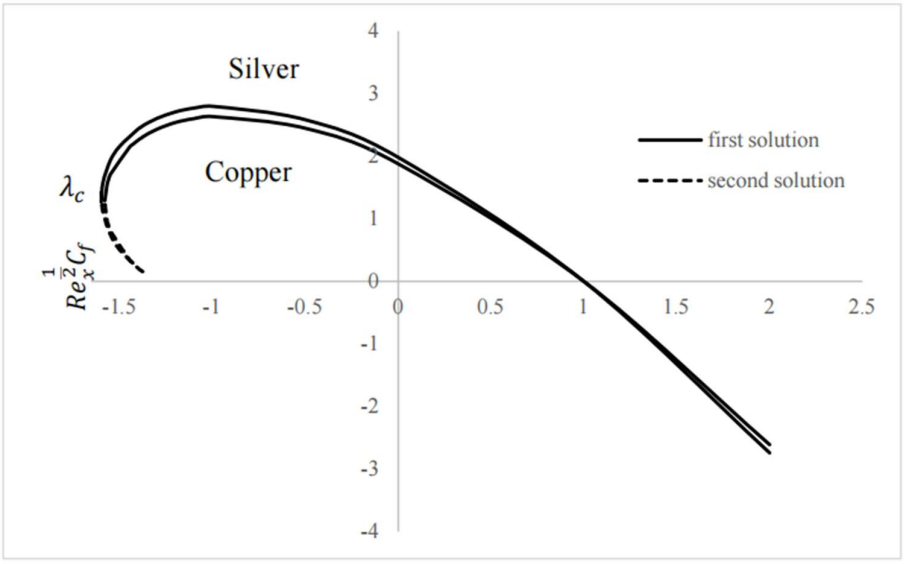


Figure 3.5: The Skin Friction Coefficient, $Re_x^{\frac{1}{2}}C_f$ versus λ with different type nanoparticle(Cu, Ag), $m = 1, \phi = 0.1, s = 0.5$.

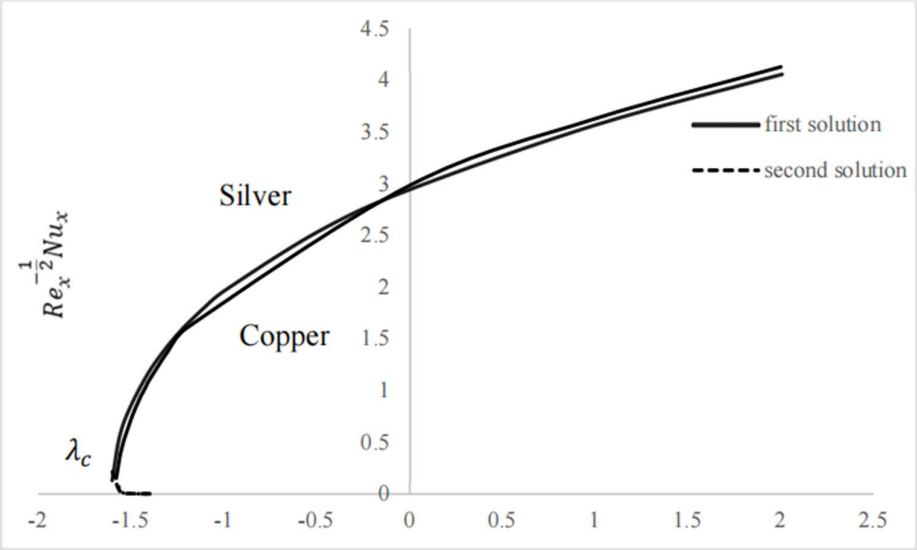


Figure 3.6: Local Nusselt number, $Re_x^{-\frac{1}{2}}Nu_x$ versus λ for different type nanoparticles (Cu, Ag) when $m = 1, \phi = 0.1\ s = 0.5$.

4. Conclusion

The purpose of this study is to investigate nanofluid flow over a permeable stretching and shrinking surface. Similarity reduction variables are used to transform the nonlinear partial differential equations (PDEs) into ordinary differential equations (ODEs). The equations are then numerically solved using MATLAB's built-in BVP4C solver. The results are compared with previous research to validate the accuracy of BVP4C. The findings confirm the validity of the results, as they had a good agreement with the earlier studies. Additionally, this study analyses the influence of stretching/shrinking parameters and the suction parameter on the velocity and temperature profiles, as well as the effect of copper and silver nanoparticles on the skin friction coefficient and local Nusselt number.

5. Acknowledgments

The authors would like to thank the Faculty of Computer Science and Mathematics, Universiti Teknologi Mara, Dr. Siti Hidayah Binti Muhad Saleh as supervisor for the support.

6. References

[1] Sreelakshmy K.R, Nair, A. S.,& Nair, S. C. (2014). Av overview of recent research. *International Research Journal Of Pharmacy*, *5*(4), 239–243. https://doi.org/10.7897/2230-8407.050451

[2] Wahid, N. S., Arifin, N. M., Khashi'ie, N. S., & Pop, I. (2023). Three-dimensional unsteady radiative hybrid nanofluid flow through a porous space over a permeable shrinking surface. *Chinese Journal of Physics*, *85*, 196–211. https://doi.org/10.1016/j.cjph.2023.07.016

[3] Zaimi, K., & Ishak, A. (2015). Boundary layer flow and heat transfer over a permeable stretching/shrinking sheet with a convective boundary condition. *Journal of Applied Fluid Mechanics*, *8*(3), 499–505. https://doi.org/10.18869/acadpub.jafm.67.222.22793

[4] Yahaya, R. I., Mustafa, M. S., Md Arifin, N., Pop, I., Md Ali, F., & Mohamed Isa, S. S. P. (2023). Hybrid nanofluid flow past a biaxial stretching/shrinking permeable surface with radiation effect: Stability analysis and heat transfer optimization. *Chinese Journal of Physics*, *85*, 402–420. https://doi.org/10.1016/j.cjph.2023.06.003

[5] Patil, P. M., Kulkarni, M., & Hiremath, P. S. (2020). Effects of surface roughness on mixed convective nanofluid flow past an exponentially stretching permeable surface. *Chinese Journal of Physics*, *64*, 203–218. https://doi.org/10.1016/j.cjph.2019.12.006

[6] Konai, S., & Mukhopadhyay, S. (2024). Nanofluid flow past a nonlinearly stretched surface with shear flow and zero nanoparticle flux. *Partial Differential Equations in Applied Mathematics*, *9*. https://doi.org/10.1016/j.padiff.2024.100643

[7] Idris, S., Jamaludin, A., Nazar, R., & Pop, I. (2023). Radiative MHD flow of hybrid nanofluid over permeable moving plate with Joule heating and thermal slip effects. *Alexandria Engineering Journal*, *83*, 222–233. https://doi.org/10.1016/j.aej.2023.09.065

[8] Patel, V. K., Pandya, J. U., & Patel, M. R. (2023). Testing the influence of TiO2− Ag/water on hybrid nanofluid MHD flow with effect of radiation and slip conditions over exponentially stretching & shrinking sheets. *Journal of Magnetism and Magnetic Materials*, *572*. https://doi.org/10.1016/j.jmmm.2023.170591

[9] Zainal, N. A., Nazar, R., Naganthran, K., & Pop, I. (2021). Unsteady MHD stagnation point flow induced by exponentially permeable stretching/shrinking sheet of hybrid nanofluid. *Engineering Science and Technology, an International Journal*, *24*(5), 1201–1210. https://doi.org/10.1016/j.jestch.2021.01.018

[10] Khan, U., Waini, I., Ishak, A., & Pop, I. (2021). Unsteady hybrid nanofluid flow over a radially permeable shrinking/stretching surface. *Journal of Molecular Liquids*, *331*. https://doi.org/10.1016/j.molliq.2021.115752

[11] Wahid, N. S., Arifin, N. M., Yahaya, R. I., Khashi'ie, N. S., & Pop, I. (2024). Impact of suction and thermal radiation on unsteady ternary hybrid nanofluid flow over a biaxial shrinking sheet. *Alexandria Engineering Journal*, *96*, 132–141. https://doi.org/10.1016/j.aej.2024.03.079

[12] Arifin, N. M., Nazar, R., & Pop, I. (2011). Viscous flow due to a permeable stretching/shrinking sheet in a nanofluid. *Sains Malaysiana*, *40*(12), 1359-1367. https://core.ac.uk/download/pdf/11492130.pdf

[13] Ismail, N. H., & Aman, F. (2023). Solving Nanofluid Flow Over a Permeable Stretching and Shrinking Surfaces using Shooting Technique. *Enhanced Knowledge in Sciences and Technology*, *3*(1), 064-069.
https://penerbit.uthm.edu.my/periodicals/index.php/ekst/article/view/10309

Default Risk Prediction using Merton's Model and Altman's Z-Score

Anas Ahmad[1] , Mohd Ezzaden Jasni [2], Mohd Shahirul Fazli Mohamad Hatta[3], Norliza Muhamad Yusof[4,*]
[1,2,3,4]*Kolej Pengajian Pengkomputeran, Informatik dan Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, Malaysia*
[*]norliza3111@uitm.edu.my

Abstract

This research focuses on predicting the default risk of a financially distressed firm with KMV-Merton's model and Altman's Z-Score model. Part of the objectives of this research were to estimate the probability of default with Merton's model and to compare and validate this probability with Altman's Z-Score. Merton's model, which is a structural model, was used to estimate default probability, while Altman's Z-Score model, a reduced-form model, was used as a verification tool. The research involved collecting past financial data from the company's financial statements. The result indicated an extremely high degree of correspondence among the models. Merton's model estimated a default probability of 100%, while Altman's Z-Score model gave a default risk forecast of -4.018, which is an exceptionally high bankruptcy likelihood. These were expected results within the known standards of both models. Challenges in research were encountered primarily in obtaining effective data and achieving a final comparability tool for model accuracy. Future research can explore variations of model comparison to further evaluate the predictive ability of Merton's model to predict default risk and review enhancements to Altman's Z-Score model to enhance prediction accuracy. The practical significance of this study lies in providing a comparative study on two of the most used default risk models and highlighting the importance of solid financial distress assessment for companies with a history of financial distress.

Keywords*:* Default Risk, Merton's Model, Altman Z-Score, Bankruptcy Prediction

Introduction

Businesses exist in various forms, including sole proprietorships, partnerships, and corporations. This study focuses on companies, which commonly prioritize profit maximization and risk minimization. Effective business management and the avoidance of bankruptcy are paramount to achieving these goals. Furthermore, [1] demonstrated proactive risk management approach ensuring firms financial stability and longevity, especially for banks [2 – 6]. Ineffective management of default risk can lead to substantial losses, financial distress, and even bankruptcy. This research seeks to contribute to a better understanding of default risk assessment to help companies mitigate these negative impacts.

Companies often face challenges in accurately assessing the probability of default when evaluating borrowers or their financial stability. This difficulty arises from the lack of universally accepted or easily applicable methods for quantifying default risk, leading to potential misjudgements in lending decisions and financial planning. While various models exist for predicting default risk, empirical tests are necessary to determine their accuracy and

reliability. This study fills this gap by comparing the performance of two widely used models, Merton's model and Altman's Z-Score model, to analyse their concordance in estimating default risk.

Merton's model provides a powerful approach to predicting firm defaults based on market value [7], while Altman's Z-Score model can forecast firms' default by analysing the firm financial status over periods longer than one or two years [8]. Some studies would focus on the theoretical nature of default risk models, while others would utilize simulated data. There may be a shortage of research that properly tests the models against actual firm performance and financial outcomes over some period. Merton's model and Altman's Z-Score performance may differ when estimating default risk in the case of a chosen firm or industry. This research can do this by employing the models on historical data and contrasting the model predictions with the actual financial trajectory of the company, thus providing real-world validation. So, this research contributes by providing a direct comparison of predictive power of these models in one specific case, empirically supporting the argument regarding their relative strengths and weaknesses.

KMV-Merton Model

KMV-Merton's model, a structural model, is a cornerstone in default probability assessment. The model emphasizes its reliance on the market value of equity as a key input for robust results [9]. The model utilizes option pricing theory, conceptualizing default as occurring when a company's asset value falls below its debt obligations at a specific time [10]. Cakici et al. [11] show that the KMV-Merton model framework can capture complex relationships between default risk and financial positions which highlights the model's practical applicability. Jumbe and Gor [12] suggest the key determinants in assessing default probabilities and show the relevance of volatility and expected returns in hedging default risk exposures using the KMV-Merton model. Specifically, [10] present a three-step method of estimating default probability, based on KMV model:1) estimation of asset value and volatility; 2) calculation of distance-to-default; and 3) calculation of default probability from distance-to-default.

The KMV-Merton model is nonetheless relevant in the modern financial markets in estimating probabilities of default [13]. A decade ago, [14] employed the Merton's model in forecasting default risk among publicly non-financial firms in the United Kingdom. The research demonstrated that the model provides a strong signal of potential failure one year in advance, showcasing its predictive power. Furthermore, [14] found that Merton's model outperformed a reduced-form model in predicting default, demonstrating its relative effectiveness. Norliza and Maheran [15] implemented Merton's model to estimate default probabilities value of Malaysian companies listed on Bursa Malaysia. The value of the default probabilities was then compared with the company rating by Malaysian Rating Corporation Berhad (MARC) and RAM Rating Holding Berhad. The research identified a high degree of correlation between the predictions made by the model and the ratings provided by the rating agencies. This indicates that Merton's model is efficient in capturing default probabilities and can accurately forecast changes in the rating outlook of companies up to two years ahead. Furthermore, it supports the application of the model in an emerging market scenario. As time passes by, the KMV-Merton model goes on to further evolve with many modifications [16-18].

Finally, the literature proves that the Merton model remains a primary method of forecasting default risk due to its consistency with theory and adaptability. Its applications span the entire

economy, as an example of continuing to be at the leading edge of innovation and application when it comes to default risk management.

Altman's Z-Score Model

Altman's Z-Score model is a widely used tool for predicting financial distress and bankruptcy [19]. Its ability to analyse a company's financial health and predict potential financial problems has been highlighted in many studies [20-21]. Altman et al. [22] developed Altman's Z-Score model for manufacturing and non-manufacturing companies. The Altman's Z-Score model has become a key input for many internal rate-based models [22]. Utilized for over many years, Altman's Z-Score model remains a fundamental tool for bankruptcy and financial problem forecasting in research and practice. Altman collected a list of 22 potentially important financial ratios from balance sheet and income statement data, categorized into liquidity, leverage, profitability, solvency, and activity ratios. Altman selected five ratios for corporate bankruptcy prediction, using statistical significance, relative responsibilities of variables, interconnections between factors, predictive accuracy, and investigator judgment. Altman et al. [22] noted that modifications to the Altman's Z-Score model require changes to financial data and estimation techniques to improve results.

Gurau [23] emphasizes that Altman's Z-Score model is a valuable tool for investors to forecast bankruptcy. In a study of 33 bankrupt US companies, [23] found that by utilizing the information on company size, financial structure, performance, and current liquidity, bankruptcy can be predicted one year in advance. Li and Rahgozar [24] developed the use of asset volatility within Altman's Z-Score model for bankruptcy prediction. Li and Rahgozar [24] reported that approximately 95% of bankrupt organizations were correctly classified as bankrupt, and roughly 80% of non-bankrupt organizations were correctly classified as not bankrupt. Therefore, asset volatility is said to be the key variable in Altman's Z-Score model for manufacturing and non-manufacturing companies.

Mantziaris [25] used Altman's Z-Score model to assess the future of an industry after a global financial crisis, examine the impact of Chinese government regulations on the real estate industry, and predict future failures of Chinese real estate companies. Mantziaris [25] also found that Altman's Z-Score model correctly predicted company classifications by financial institutions in Lebanese manufacturing firms and company failures for the economy of Pakistan. Not only that, Altman’s model also can assess the current and future health of small and medium enterprises in the study of [26]. Singh and Singla [27] further extended the model's relevance in emerging markets such as the Indian corporate sector. Under dynamic market conditions, the Altman Z-Score model shows robustness in exploring the impact of environmental considerations on companies' financial health [28]. In terms of comparative analyses, the reliability of the Altman Z-Score model remains satisfactory compared to the alternative model [29].

In summary, the Altman Z-score model is still a great predictor of corporate default, as reflected by the diverse applications by industries and nations. Its empirical resilience, flexibility, and age-old prescience prove the model's strategic location in financial distress diagnosis and corporate governance.

Data Description and Sources

The selection of appropriate financial distress prediction models and the acquisition of relevant financial data are necessary to carry out an elaborate analysis for firms with a past record of financial distress. The study of an undisclosed Malaysian public manufacturing company firm characterized by a history tainted with massive financial losses, is an appropriate subject for financial distress analysis. Given this history of financial distress, models designed to assess the likelihood of bankruptcy, such as Merton's and Altman's Z-Score, are particularly suitable for analysing the company under study. By obtaining the company-specific financial data and implementing these models, this study can perform a sound and evidence-supported analysis of the company's financial distress.

Data for the selected Malaysian undisclosed public manufacturing company was gathered from two sources: the company's 2015 annual report and Bursa Malaysia's archived stock price database. The annual report gave financial variables including the number of outstanding shares, short-term borrowings, total liabilities, total assets, current assets, current liabilities, earnings before taxation, earnings after taxation, revenue, and book value of equity. Corresponding monthly stock prices were fetched from Bursa Malaysia website.

Applying the Merton's model

Merton's model, a framework for assessing a firm's default risk, necessitates information on outstanding shares, liabilities, and stock prices. In this study, these data were used to compute the following variables:

a. The market value of equity, E: Calculated by multiplying the monthly stock price with the fixed number of outstanding shares issued by the company for one year, generating twelve monthly market values of equity starting for the year 2015.
b. Book value of debt, k: Assumed to be equivalent to the value of short-term borrowings reported in the annual report, and it is treated as constant.
c. Market Value of Asset, V: Derived by summing up the monthly values of E with the fixed value of k.
d. Growth or drift rate, μ: Defined as the firm's return on assets (ROA). Monthly ROA was calculated by dividing net income (fixed value) by the monthly values of V. Then, the annual growth rate was obtained by averaging these monthly ROA values.
e. Volatility of Asset, σ: Calculated as the annualized standard deviation of the monthly returns of V, achieved by multiplying the standard deviation by the square root of 12.

Based on the computed variables, the probability of default of the company is estimated. In this context, probability of default, indicates a probability (Pr) that the the market value of the company asset at time t, V_t will be worth less than the book value of liabilities, k by the time the debt matures [9 -10]. This is expressed as follows:

$$P_t = \Pr[V_t \leq k] \tag{1}$$

. Let V_t be a function that satisfies the following stochastic differential equation:

$$V_t = \ln V + (\mu - \frac{\sigma^2}{2})t + \sigma\sqrt{t}\varepsilon \tag{2}$$

where ln V is defined as a log-distributed of an initial value market value of asset, μ is the drift rate, σ is the asset volatility, and ε is the random component of the firm's asset returns. Substituting (2) into (1):

$$P_t = \Pr\left[\ln V + (\mu - \frac{\sigma^2}{2})t + \sigma\sqrt{t}\varepsilon \le \ln k\right] \tag{3}$$

Then, rearranging (3)to be as follows:

$$P_t = \Pr\left[-\frac{\ln\left(\frac{V}{k}\right) + \left(\mu - \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}} \ge \varepsilon\right] \tag{4}$$

Assuming the random component of the asset return is normally distributed $\varepsilon \sim N(0,1)$. Hence, the probability of default in terms of the standard cumulative normal distribution can be written as

$$P_t = N[-d] \tag{5}$$

where d is the distance-to-default of the firm that is expressed as:

$$d = \frac{\ln\left(\frac{V}{k}\right) + \left(\mu - \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}} \tag{6}$$

The probability of default and distance to default of the company were calculated using equations 5 and 6, respectively with t =1 year.

Applying the Altman's discriminant model

This study also employed Altman's Z-Score model, a multivariate financial distress prediction tool that utilizes five distinct financial ratios (X_1, X_2, X_3, X_4, and X_5). The model aims to classify companies based on their likelihood of financial distress, with lower Z-scores indicating higher odds of bankruptcy. For public manufacturing companies, Altman modified the model using the book value of equity in place of market capitalization and applying specific variable weights. The data collected was utilized in this revised Altman's Z-Score model calculated as below [22]:

$$Z = 0.717X_1 + 0.847X_2 + 3.107X_3 + 0.420X_4 + 0.998X_5 \tag{7}$$

where,
X_1: Working Capital (current assets - current liabilities) / Total Assets
X_2: Retained earnings (Earnings after taxation) / Total Assets
X_3: Earnings Before Interest and Taxes / Total Assets
X_4: Book Value of Equity / Total Liabilities,
X_5: Sales (revenue) / Total Assets.

Comparative Analysis of Merton's and Altman's Z-Score Models

A comparative analysis was conducted between the default risk assessments provided by Merton's and Altman's Z-Score models. The comparison was made based on Tables 1 and 2.

Table 1: Default Rates Classification

Default rate %	Classification
0.01-1.41	Firms with a very strong capacity to meet financial commitments. Low default risk. A bond rated as AAA to A-
2.30-32.50	Firms with an adequate capacity to meet financial commitments. Moderate default risk. Bond rated as BBB to B-.
46.61-100	Firms with immediate financial problems, indicate a high probability of default. High default risk. A bond rated as CCC to D.

According to Table 1 [30], Merton's model ratings were categorized into three risk levels: low, moderate, and high default risk. Each level is represented by rating and default rate thresholds. Meanwhile, Altman's Z-Score model classified companies into three categories: non-bankruptcy, gray zone, and bankruptcy, based on the thresholds shown in Table 2 [31].

Table 1: Altman's Z-Score Model Rating for Public Manufacturing Companies

Z-Score Range	Classification
Above 2.9	Non- bankruptcy
1.23 – 2.9	Gray Zone
Below 1.23	Bankruptcy

Results and Discussion

The default risk assessments from Merton's model are presented and discussed based on the data in Figure 1 and Table 3. Figure 1 illustrates the market value of assets trend for an undisclosed company. The graph clearly shows significant fluctuations in the market value of assets throughout the year. This indicates volatility in the company's financial performance. The market value of assets is at its lowest point in January, indicating a potentially challenging start to the year. Then, the market value of assets reaches its highest point in July, suggesting a period of relatively strong performance or investor confidence during that month. After the peak in July, there is a general downward trend in the market value of assets, suggesting a potential decline in the company's financial health during the latter part of the year.

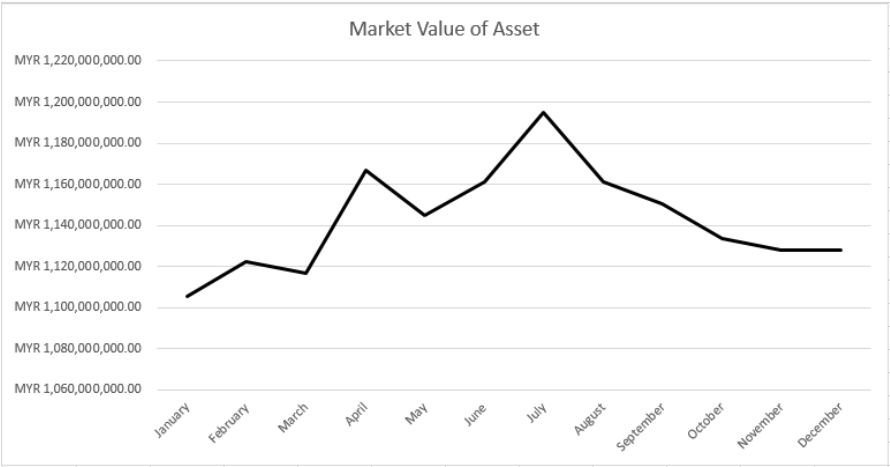


Figure 1: Market Value of Assets Trend for an Undisclosed Company in 2015

The volatility depicted in the graph supports the volatility figure (4.595%) calculated in Merton's model (Table 3). This volatility contributes to the negative distance to default and the 100% probability of default. The growth rate of -60.192% suggests significant losses. This further reinforces the company's precarious financial position, indicating that the company was deemed unable to meet its debt obligations.

Table 3: Merton's Model Default Risk Metrics

Volatility, σ	Growth Rate, μ	Distance-to-default, d	Probability of default, P_t
4.595%	-60.192%	-12.015	100%

Table 4 presents the annualized coefficients of Altman's Z-Score model for 2015. The calculated Z-Score of -4.018 indicates that the company was in the bankruptcy category, according to the criteria outlined in Table 2.

Table 4: Altman's Z-Score Coefficients

2015	x_1	x_2	x_3	x_4	x_5	Z
Value	-1.8888	-0.6767	-0.6768	0.0281	0.0008	-4.018

The overall results suggest that the company was facing financial challenges throughout 2015, which aligns with the conclusions drawn from both Merton's and Altman's models as concluded in Table 5. Table 5 compares the default risk assessments from Merton's and Altman's models. Both models indicate a high risk of bankruptcy. Merton's model predicts a 100% probability of default, while Altman's Z-Score indicates bankruptcy with a score of -4.018. These findings align with the company's documented financial difficulties, as discussed in the data section. The consistency between the results of Merton's and Altman's models supports the assessment of the company's financial distress.

Table 5: Comparison of Merton's and Altman's Model Results

Model	Merton's	Altman's Z-Score
Likelihood to default	100%	-4.018
Classification	High Default Risk	Bankruptcy

Conclusion

This research aimed to assess the default risk of a company using Merton's model and Altman's Z-Score model. Merton's model was employed to calculate the probability of default, while Altman's Z-Score model was employed to validate the estimated probability of default. These two models were employed to test the default risk of the selected undisclosed Malaysian manufacturing public company. Furthermore, this study provided a methodology for calculating key inputs for Merton's model, including the market value of assets, and asset volatility. Consequently, the distance-to-default was also determined. The results obtained from both models indicated high levels of agreement. Merton's model gave a default probability of 100%, while Altman's Z-Score model offered a default risk of -4.018. These results indicate a high likelihood of bankruptcy. The results are consistent with established benchmarks. The 100% default rate predicted by Merton's model aligns with the characteristics of a company in a state of financial distress. Similarly, Altman's Z-Score result of -4.018 falls within the range associated with bankruptcy.

This study encountered limitations, primarily in obtaining precise data corresponding to the annual reports of the selected company. Additionally, there was a challenge in establishing a definitive tool to measure the accuracy of the model comparison as default rates can be varied across companies. Future research could explore variations in model comparisons to further evaluate the effectiveness of Merton's model in forecasting default risk. In the context of Altman's Z-Score model, future studies might investigate the potential for refining the equation by adding or removing parameters to enhance default risk prediction accuracy.

Acknowledgments

This research paper would not have been possible without the support of facilities and resources from the UiTM Cawangan Negeri Sembilan, Kampus Seremban. Also, our sincere gratitude to the editorial team's and reviewers' contributions, which strengthened the clarity and consistency of this research paper.

References

[1] Du, Y. & He, Y. (2024). Deleveraging, debt default risk, and enterprise innovation: Evidence from China. *Transactions on Economics, Business and Management Research*, 6, 250-263. https://doi.org/10.62051/ean14687

[2] Akomeah, J., Agumeh, R., & Frimpong, S. (2020). Credit risk management and financial performance of listed banks in Ghana. *Research Journal of Finance and Accounting*, 11, 6, 39-48. https://doi.org/10.7176/rjfa/11-6-05

[3] Chen, J. & Ma, L. (2022). Bank credit risk avoidance and countermeasures based on wireless communication. *Scientific Programming*, 1-7. https://doi.org/10.1155/2022/7793088

[4] Tien, D. P. & Nguyen, K. Q. T. (2023). Credit risk management and effectiveness of credit activities: Evidence from Vietnam. *Journal of Eastern European and Central Asian Research* (JEECAR), 10(5), 788-799. https://doi.org/10.15549/jeecar.v10i5.989

[5] Sutono, J. & Hermawan, A. A. (2024). Implementation of credit risk management Basel principles (evaluation study of an Indonesian commercial bank). *Owner*, 8(2), 1190-1199. https://doi.org/10.33395/owner.v8i2.2045

[6] Kamara, A. K. (2024). The study of credit risk in the banking sector and its effect on financial performance case study of the zenith bank Sierra Leone. *European Journal of Economic and Financial Research*, 8(4). https://doi.org/10.46827/ejefr.v8i4.1732

[7] Christoffersen, B., Lando, D., & Nielsen, S. F. (2022). Estimating volatility in the Merton model: The KMV estimate is not maximum likelihood. *Mathematical Finance*, 32(4), 1214-1230. https://doi.org/10.1111/mafi.12362

[8] Altman, E. I. (2000). Predicting financial distress of companies: Revisiting The Z-score and Zeta Models. *Journal of Banking and Finance*, 1(2); 1968 – 2000.

[9] Hull, J., Nelken, I., & White, A. (2004). Merton's Model, Credit Risk, and Volatility Skews. *Journal of Credit Risk*, *1*(1), 1–17.

[10] Crosbie, P., & Bohn, J. (2003). Modeling Default Risk - Modeling Methodology. *Moody's KMV Working Paper*, 1–31. http://doi.org/10.1016/j.frl.2006.03.003

[11] Cakici, N., Chatterjee, S., & Chen, R. (2019). Default risk and cross section of returns. *Journal of Risk and Financial Management*, 12(2), 95. https://doi.org/10.3390/jrfm12020095

[12] Jumbe, G.& Gor, R. (2023). Examining the impact of debt maturity time, expected return and volatility on probability of default in credit risk modelling: The case of Merton and MKMV models. *International Journal of Engineering Science Technologies*, 7(1). https://doi.org/10.29121/ijoest.v7.i1.2023.442

[13] Aminda, R. S., Aziz, M. J., Suharti, T., subakti, J., Wibowo, S. S., & Subakti, A. K. (2023). Determining credit risk using contingent claim model approach (Merton model): A case study of Indonesian digital banks. *Advances in Economics, Business and Management Research*, 74-79. https://doi.org/10.2991/978-94-6463-244-6_14

[14] Tudela, M., & Young, G. (2005). A Merton-model approach to assessing the default risk of UK public companies. *International Journal of Theoretical and Applied Finance*, 8(06), 737–761.

[15] Norliza, M. Y., & Maheran, M. J. (2012). Estimating the Credit Risk of Malaysian Companies using Merton Model. World Applied sciences Journal, 17(72-78), 1-7.

[16] Miao, H., Ramchander, S., Ryan, P., & Wang, T. (2018). Default prediction models: The role of forward-looking measures of returns and volatility. *Journal of Empirical Finance*, 46, 146-162. https://doi.org/10.1016/j.jempfin.2018.01.001

[17] Yusof, N. M., Alias, I. Q., Kassim, A. J. M., & Zaidi, F. L. N. M. (2021). Determining the credit score and credit rating of firms using the combination of KMV-Merton model and financial ratios. *Science and Technology Indonesia*, 6(3), 105-112. https://doi.org/10.26554/sti.2021.6.3.105-112

[18] Oluwo, A. K. and & Villamor, E. (2022). Quantitative structural models to assess credit risk on individuals. *Journal of Applied Mathematics and Physics*, 10(07), 2313-2340. https://doi.org/10.4236/jamp.2022.107158

[19] Altman, E. I. (2018). Applications of distress prediction models: What have we learned after 50 years from the z-score models? *International Journal of Financial Studies*, 6(3), 70. https://doi.org/10.3390/ijfs6030070

[20] Hayes, S. K., Hodge, K. A., & Hughes, L. W. (2010). A Study of the Efficacy of Altman's Z to Predict Bankruptcy of Specialty Retail Firm Doing Business in Contemporary Times. *Economics & Business Journal: Inquiries & Perspectives,* 3(1), 1-13.

[21] Gabbi, G. & Levich, R. M. (2019). Controlling risks to ensure financial stability and reducing volatility. *Journal of International Financial Management & Amp Accounting*, 30(3), 183-187. https://doi.org/10.1111/jifm.12107

[22] Altman, E. I., Drozdowska, M. I., Laitinen, E. K., & Suvas, A. (2014). Distressed firm and bankruptcy prediction in an international context: A review and empirical analysis of Altman's Z-Score Model (August 10, 2014). *Available at SSRN: https://ssrn.com/abstract=2536340* or http://dx.doi.org/10.2139/ssrn.2536340

[23] Gurau, T. (2013). A model of bankruptcy prediction. Business Economics. Bachelor Thesis, Erasmus University of Rotterdam. Retrieved from http://hdl.handle.net/2105/13759

[24] Li, J., & Rahgozar, R. (2012). Application of the Z-Score Model with Consideration of Total Assets Volatility in Predicting Corporate Financial Failures from 2000-2010. *Journal of Accounting and Finance,* 12 (2), 1-9.

[25] Mantziaris, S. Z. (2015). Bankruptcy Prediction Models: An Empirical Analysis of Altman's Z-Score model in Forty Greek Companies in the Period of Economic Recession. Master Thesis, University of Macedonia, Greece.

[26] Mitrašević, M. & Bardarova, S. (2020). Measuring the risk of lending to small and medium-sized enterprises in the republic of Serbia in light of modern banking regulations. *Ekonomski Horizonti*, 22(3), 263-277. https://doi.org/10.5937/ekonhor2003263m

[27] Singh, G. & Singla, R. (2021). Default risk and stock returns: Evidence from Indian corporate sector. *Vision: The Journal of Business Perspective*, 27(3), 347-359. https://doi.org/10.1177/09722629211003358

[28] Oyegunle, A., Weber, O., & Elalfy, A. F. (2023). Carbon costs and credit risk in a resource-based economy: carbon cost impact on the z-score of Canadian TSX 260 companies. *Journal of Management and Sustainability*, 13(1), 187. https://doi.org/10.5539/jms.v13n1p187

[29] Rybárová, D., Majdúchová, H., Štetka, P., & Luščíková, D. (2021). Reliability and accuracy of alternative default prediction models: Evidence from slovakia. *International Journal of Financial Studies*, 9(4), 65. https://doi.org/10.3390/ijfs9040065

[30] Altman, E.I. & Kishore, V. (2000). The Default Experience of U.S. Bonds, Working Paper, Salomon Center, New York University.

[31] Kumar, M. N., & Rao, V. S. H. (2014). A New Methodology for Estimating Internal Credit Risk and Bankruptcy Prediction under Basel II Regime. *Computational Economics, 46*(1), 83-102. doi: 10.1007/s10614-014-9452-9