J E M U R

UITMCNS SEREMBAN

JOURNAL OF EXPLORATORY MATHEMATICAL UNDERGRADUATE RESEARCH

RECENT **MATHEMATICAL PROJECTS** AT UITM, *SEREMBAN CAMPUS*

Vol. 002 | PREPARED BY: KPPIM, UiTM N. SEMBILAN. | DATE: 05 OCT 2024

Photography courtesy of Porapak Apichodilok

JOURNAL OF
EXPLORATORY MATHEMATICAL
UNDERGRADUATE RESEARCH

OFFICIAL JOURNAL OF THE COLLEGE OF COMPUTING, INFORMATICS, AND MATHEMATICS, UITM NEGERI SEMBILAN BRANCH, SEREMBAN CAMPUS, OCTOBER 2024 EDITION

DISCLAIMER:

All articles published in the JEMUR 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.

UITMCNS
SEREMBAN

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

eISSN: 3030-5411
Perpustakaan Negara Malaysia
This journal is published twice a year

JOURNAL OF
EXPLORATORY
MATHEMATICAL
UNDERGRADUATE RESEARCH

Foreword from the Editor in Chief

We express our gratitude to Allah SWT for the publication of the second edition of the Journal of Exploratory Mathematical Undergraduate Research (JEMUR). This journal is published by the College of Computing, Informatics, and Mathematics (KPPIM) at UiTM Seremban campus.

This journal seeks to showcase the research discoveries of students and lecturers engaged in the Final Year Projects of the Bachelor's Degree Programme at the KPPIM, UiTM Seremban campus.

Starting from 2014, final year students from KPPIM (formerly called FSKM) at the UiTM Seremban campus have made notable contributions to academic research. However, many projects have had challenges in getting their work published in academic conference proceedings or journals due to specific hurdles, such as the level of scientific research and writing quality. We expect that the publication will facilitate the dissemination of KPPIM's Seremban research findings through other channels.

This publication serves as a platform for disseminating the most recent research conducted by students and lecturers from KPPIM Seremban.

As a representative of the editorial team, I extend my compliments and commendations to the publishing team for their diligent efforts in producing the JEMUR papers this time.

Thank you.

JOURNAL OF EXPLORATORY MATHEMATICAL UNDERGRADUATE RESEARCH
Volume 2 (October 2024)

College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus

Contents

PREFERABLE HIGHER EDUCATIONAL INSTITUTIONS BY MATRICULATION STUDENTS USING ANALYTICAL HIERARCHY PROCESS (AHP)

Rasidah Buang[1], Maznah Banu Habiboo Raman[2], Rahmah Shahril[3], Wardina Syafiah Mohd Fadzil[4], Nurul Falah Mohd Razali[5], Nurul Huda Shahrudin[6]

[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Cawangan Melaka, Kampus Jasin, Melaka. [2] College of [2,3,4,5,6]Computing,Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan

Corresponding author: rasid396@uitm.edu.my

Abstract

The students must be aware of their options for higher education study and comprehend their options for study pathways before they apply to a university. Choosing the right path is very important for students and it will be a difficult process if the students are unprepared. Poor preparation and knowledge about degree planning will lead to confusion among students due to the variety of higher-level institution options available to a student. This study will identify the factors that influence students' decisions to attend higher education institutions using the Analytical Hierarchy Process (AHP)and the preferred institutions that students in this cohort pick and examine the nature of the connection between the institutions that students prefer. The main sample for this study was drawn from Melaka's Matriculation College students over the age of 18. The data is gathered by analyzing the preferred university selection process and includes the identification of relevant criteria that are found necessary by students, such as college fees, friends and family influence, career path availability, course availability, the location of institutions, and scholarship coverage. The finding reveals that most students chose a public institution (IPTA) based on the availability of required courses and programs. This study is expected to be significant for professionals, higher education institution managements, and future researchers

***Keywords**: AHP; Analytic Hierarchy Process;higher educational institution; criteria; Multicriteria Decision Making*

1. Introduction

Higher education leads to the award of a degree. It is also known as post-secondary education and it is an optional last stage of formal learning for any students who desire to continue their education after finishing for a certificate which is Sijil Matrikulasi KPM. The idea of attending a university to study is exciting to many students. They will experience a new environment, meet new people, and gain new knowledge. However, students must be aware of their options for higher education study and comprehend their options for study pathways before they apply to a university. Choosing the right path is very important for students, and it will be a difficult process if the students are unprepared. Matriculation program is a pre-university education that typically lasts between two and four semesters and, upon successful completion, permits students to apply for degree programs at public universities (Abdussyukur et al., 2021). Some matriculation students lack a structured application to assist them in planning their degree based on certain factors. Hence poor preparation and knowledge about degree planning will lead to confusion among students due to the variety of higher-level institution options available to a student.

The Analytic Hierarchy Process (AHP) is a methodical approach for expressing any problem's elements hierarchically. In 1971, Thomas and Saaty (1990) created the AHP, as a method for methodical decision-making. It is a system of organizing decision-making mechanisms in a scenario influenced by several different independent actors (Kuzu, O. H., 2020). They provide a method for decision-making when there are limited choices, but each has a variety of attributes. It is also used in decision making situations where multiple factors must be considered simultaneously. Principle eigenvectors are used to create the ratio scale whereas principal eigenvectors are to create the consistency index (Lee, 2016). In this study will identify the factors that influence students' decisions to attend higher education institutions using the Analytical Hierarchy Process (AHP). It is also will determine the preferred institutions that students in this cohort pick and examine the nature of the connection between the institutions that students prefer; and the selection criteria that influence their decisions. This study deals with the application of the Analytical Hierarchy Process.

The objective of this study is to identify the factors of higher education institutes selection by matriculation students using Analytic Hierarchy Process (AHP) and to determine the preferable institutes of higher education chosen by matriculation students using Analytic Hierarchy Process (AHP)

2. Methodology

There are 5 steps in AHP approach which are identifying criteria and alternative, constructing a hierarchy framework analyzing data, collecting information by using questionnaires and calculating the weightage for criteria and alternative. The criteria that have been considered in this study are amount of college fees, friends and family influences, availability of career paths, availability of required courses or programs, location of institutions, and availability of scholarship coverages. Meanwhile 4 institutions: Public Institution (IPTA), Private Institution (IPTS), Polytechnic Premiere (Poly) and Institution of Teacher Education (IPG) have been selected as alternatives for this study.

Step I : Construct the pairwise comparison matrix

Each criterion is compared with every other criterion using the fundamental scale that gives the relative importance value shown in Table 1. Saaty,T .L (1977)

Table 1: The Fundamental Scale

Importance Scale	Definition of Importance Scale
1	Extremely importance
3	Very importance
5	Strong importance
7	Moderate importance
9	Equal importance
2, 4, 6, 8	Intermediate value

A pairwise comparison matrix is constructed in the form:

$$C = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14}\ldots\ldots & C_{1n} \\ C_{21} & C_{22} & C_{23} & C_{24}\ldots\ldots & C_{2n} \\ C_{31} & C_{32} & C_{33} & C_{34}\ldots\ldots & C_{3n} \\ \cdots \\ C_{n1} & C_{n2} & C_{n3} & C_{n4}\ldots\ldots & C_{nn} \end{bmatrix} \tag{1}$$

where C_{ij} represents the relative importance of criterion *I* over criterion *j* and *n* is the number of criteria.

Step II: determine the normalized matrix.

$$C_N = \begin{bmatrix} V_{11} & V_{12} & V_{13} & V_{14}\ldots\ldots & V_{1n} \\ V_{21} & V_{22} & V_{23} & V_{24}\ldots\ldots & V_{2n} \\ V_{31} & V_{32} & V_{33} & V_{34}\ldots\ldots & V_{3n} \\ \cdots \\ V_{n1} & V_{n2} & V_{n3} & V_{n4}\ldots\ldots & V_{nn} \end{bmatrix} \tag{2}$$

Vij is element of normalized matrix
Where

$$V_{ij} = \frac{C_{ij}}{C_j}$$

Step III: calculation of weightage

$$W_i = \frac{1}{n}\sum_{1}^{n} V_{ij} \tag{3}$$

Where *Wi* = the weightage of criteria *i*

n is a number of criteria.

Hence, the weightage of criteria can be represented as follows:

$$W_c = [W_1 \quad W_2 \quad W_3 \quad . \quad . \quad . \quad W_n] \tag{4}$$

The results of this section can be concluded that the most important criterion is the criteria with the highest weightage.

Step IV: Checking of consistency matrix

The comparison matrix is consistence if and only if

$$CW^T = nW^T$$

However, if the equation is not satisfied, indicating inconsistency, the next step is calculating the consistency index (CI) and consistency ratio (CR). The comparison matrix is considered consistent if the CR is less than 0.1. The formula for calculating CI and CR is:

Firstly, find the maximum eigenvalue (λ_{max}):

$$\lambda_{max} = \frac{1}{n} \sum_1^n \frac{C^i W_i^T}{W_i^T} \tag{5}$$

where, C_i represents the pairwise comparison of criterion i, W_i^T is the weightage of criterion iin transpose and *n* is the total number of criteria.

Next, find the consistency index (CI) and consistency ratio (*CR*):

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

$$CR = \frac{CI}{RI}$$

where λ_{max} represents the maximum eigenvalue of the matrix, *n* is the total number of criteria, and *RI* is the random index obtained from Table 2.

Table 2: Value of Random Index (RI)

n	1	2	3	4	5	6	7	8	9	10
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

Step V: Calculate the weightage of alternative

The weight of each alternative for each criterion is calculated using the same method as in *Step I* until *Step IV*. The weightage of each alternative is multiplied by the weightage of the corresponding criteria, and the results are summed to obtain the overall weightage of each alternative. The calculation of alternative weighting(W_A) is as below:

$$= [A1 \quad A2 \quad \cdots \quad A][A] \tag{6}$$

3. Result and Discussion

Based on the research findings, the preferences of UiTM students for higher educational selection were determined.

Table 3: Final ranking for each criterion

C*n*	Criterion	Weightage	Ranking
C_1	Amount of College Fees	0.18037	4
C_2	Friends and Family Influences	0.08183	6
C_3	Availability of Career Paths	0.20942	2
C_4	Availability of Required Courses or Programs	0.21421	1
C_5	Location of Institutions	0.13001	5
C_6	Availability of Scholarship Coverages	0.18416	3

Based on Table 3, the most important factor for matriculation students to select the higher educational is availability of required course or program. Followed by availability of career paths, availability of scholarship covereage and fees. Location and influence from family and friends are the least important for them to select the higher educational institutions.

Table 4: Final ranking for each alternative

A*n*	Alternative	Weightage	Ranking
A_1	Public Institution (IPTA),	0.3928	1
A_2	Private Institution (IPTS)	0.2372	2
A_3	Polytechnic Premiere (Poly)	0.1666	4
A_4	Institution of Teacher Education (IPG)	0.2032	3

Meanwhile, in terms of preferred higher educational institution, IPTA was the top choice among students with 0.3928, followed by IPTS, 0.2372 and IPG, 0.2032. Polytechnic Premier is not the most preferrable higher educational institution among matriculation students with the weightage 0.1666.

4. Conclusion

The present study used the AHP method as a tool for evaluating the significance of the criteria that students prefer when choosing a higher education institution. The questionnaire based on AHP methodology was made. The criteria that were listed in the questionnaire are, amount of college fees, friends and family influences, availability of career paths, availability of required courses or programs, location of institutes, and availability of scholarship coverages. While the higher education institutions that were listed in the questionnaire are, Public Institution (IPTA), Private Institutions (IPTS), Institute of Teacher Education (IPG), and Polytechnic (Poly).

References

Abdussyukur, N. F., Saat, R. M., & Alias, N. (2021). Teaching and learning practices in chemistry practical work of Malaysian matriculation programme: a needs analysis. MOJES: Malaysian Online Journal of Educational Sciences, 9(4), 13-26.

Lee, S. (2016, April). Application of AHP and fuzzy AHP to decision-making problems in construction. In 52nd ASC annual international conference proceedings (pp. 13-16).

Kuzu, Ö. H. (2020). Strategy Selection In The Universities Via Fuzzy AHP Method: A Case Study. *International Journal Of Higher Education, 9(2)*, 107. https://doi.org/10.5430/ijhe.v9n2p107

Saaty, T.L (1977) A Scalling Method For Priorities In Hierarchical Structures. *Journal Of Mathematical Psycology*, *15(3)*, 234-281. https://doi.org/10.1016/0022-2496(77)90033-5.

Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. European journal of operational research, 48(1), 9-26.

CAFFEINE CONSUMPTION: ARE YOU AWARE?

Che Norhalila Che Mohamed[1], Nora Mohd Basir[2,*], Muhammad Luqman Zulkifli[3], Muhammad Shauqi Rahim[4], Najmuddin Haqimi Nazli Suhardi[5]

[1,2,3,4,5]College of Computing, Informatics and Mathematics,
Universiti Teknolgi MARA, Cawangan Negeri Sembilan, Kampus
Seremban, 70300 Seremban Negeri Sembilan, Malaysia
*noram661@uitm.edu.my

Abstract

Caffeine is found naturally in certain plants, such as coffee, tea, soft drinks, and energy beverages. Its effects, including increased heart rate, faster breathing, and heightened alertness, typically start within five to thirty minutes and vary in duration from person to person. This study aims to explore the awareness of caffeine's effects among students and the factors influencing this awareness. 397 randomly selected public university students in Negeri Sembilan participated in an email questionnaire. The analysis results indicate a notable association between daily caffeine consumption and gender. Additionally, 69.5% of the respondents are aware of the consequences of caffeine consumption. The findings from the logistic regression model demonstrate that knowledge and consumption significantly impact the level of awareness regarding the effects of caffeine consumption among students. It is essential to increase awareness among students and the community to ensure that all caffeine lovers can consume it more healthily.

Keywords: Caffeine, awareness, student, logistic.

1. Introduction

Caffeine beverage is not an unusual term among people nowadays. Caffeine is synonymous with any coffee product, whether it is a branded product like Starbucks and Coffee Bean and Tea Leave or an instant coffee product like Nescafe and Old Town White Coffee. However, caffeine cannot only be found in coffee drinks but also in other types of beverages. Caffeine is a stimulant in many plants, including coffee beans, tea leaves, and cacao beans. It is a central nervous system stimulant that can improve alertness and reduce fatigue. Caffeine can also be defined as a beverage containing stimulant drugs, widely used as a mental and physical stimulator (Aslam et al., 2013), or also known as an energy booster. Not many people acknowledge that other than coffee, beverages like tea, chocolate, and energy drinks, which are loved by people of many different levels of ages, also contain caffeine. Although those drinks do not have as much caffeine as coffee does, they still have those stimulant drugs (Aslam et al., 2013).

In today's market, many cafes supply caffeinated beverages, especially coffee, as their main beverage. Coffee shops and cafes have become popular meeting locations, and business coffee breaks are prevalent. Drinking coffee or tea can also create a sense of familiarity or comfort, making it a popular choice for people looking for familiarity or comfort. The local entrepreneur sees this business as an opportunity for a high income, as many people like coffee. It is already a trend and a lifestyle among people to drink coffee in their leisure time. Most people who consume caffeine are between 19 to 24 years old (Khan, 2019). This situation can be explained by the lifestyle of the youngster nowadays. Some do not love coffee but only want to show other people on social media to make an aesthetic post about drinking coffee in a fancy cafe.

Nevertheless, some youngsters need caffeine to do much work or study. However, teenagers from 19 to 24 years old and senior citizens over 60 are also not an exception when drinking coffee (Khan, 2019). However, these age levels are not pretending or need it as help for work. They are more likely to enjoy the taste of the coffee's uniqueness while chatting with their loved ones and friends (Pandey et al., 2021).

The after-effect of caffeine is that people can wake up for a long time, although sleepy, before consuming it. Caffeine’s potential to improve mental and physical performance is one of its most significant advantages (Astorino & Roberson, 2010). Caffeine has been shown to boost cognitive performance, including memory, attention, and reaction time. It can also improve endurance and athletic performance. People who consume caffeine also have extra energy to use to do energy-consuming work after drinking caffeine. This energy booster is an advantage to human beings. People can do many jobs past their energy limit, consuming high productivity for someone who works. Caffeine can also boost mood and lessen fatigue. Caffeine boosts the release of dopamine, a neurotransmitter associated with pleasure and reward, which can make people feel more alert and energised.

Although caffeine is such a good thing as it brings advantages to people, it also brings disadvantages. Caffeine will be a huge problem for human health if it is consumed too frequently. Numerous studies, like the study by Rodda et al. (2020) and Aslam et al. (2013), have been conducted to investigate the relationship between caffeine consumption and various health outcomes, such as cardiovascular disease, cancer, and pregnancy outcomes. While some studies have found that caffeine benefits these outcomes, others have found a negative or no effect. A study by Rodda et al. (2020) has proved that caffeine brings many health issues, such as chronic headaches and an increase in blood pressure, sleep difficulties, anxiety, and irritability. Other signs of caffeine intoxication include gastrointestinal problems, shaky hands, and tachycardia (Rodda et al., 2020).

For the past few decades, the health issue brought on by the overdose of caffeine consumption has increased rapidly. An increase in systolic and diastolic blood pressure is most likely for those who overconsume caffeine (Celi et al., 2022). There are ways to solve the over-consuming caffeine issue, such as reducing caffeine intake or finding an acceptable caffeine replacement (Rodda et al., 2020). It will decrease the possibility of getting health problems. However, it is easy to say without any action. Caffeine addiction is a severe problem that cannot be repaired in just one day. When people drink caffeine as their daily routine, they cannot easily throw that habit away in the blink of an eye. For instance, a study by Juliano et al. (2012) showed that a treatment-seeking for caffeine dependence reported that 88% of participants had at least one failed serious attempt to reduce their daily use.

Consequently, this study was conducted to identify students who were aware of caffeine's side effects. The findings would benefit the researcher by revealing the level of awareness regarding caffeine consumption and giving information about the side effects of caffeine to the community. Thus, hitting two birds with one stone will benefit society by decreasing the percentage of health issues brought on by caffeine.

2. Methodology

This study used a descriptive research design. Students from a public university in Negeri Sembilan were the target population of 5594 students. Simple random sampling was used since the sampling frame of the students was available from the university’s administrative. Using the formula from Cochran (1977) the sample size needed in this study was 397 students. The questionnaire was distributed by email to the selected students.

The questionnaire was constructed in five sections. Section A aimed to gather the respondents’ demographic information such as gender and level of study; Section B assesses the level of caffeine intake; Section C focused on the effects of caffeine consumption; Section D measured knowledge about the effects of caffeine; and Section E evaluated awareness of caffeine consumption effect. The questionnaire was adopted from the study by Ginting et al. (2022) and Arul Prakasam et al. (2022). The questionnaire in Sections A and B was measured using categorical variables, while Sections C to E were measured on a 5-likert scale.

The data analysis begins with the Chi-Square test of independence. This test determines the association between categories of daily caffeine drinkers and gender, as both variables are categorical. A crosstabulation table described the students' characteristics using descriptive statistics. A bar chart, a graphical technique, visualised the students' level of awareness of caffeine consumption. For data interpretation, mean values of 2.50 and above indicated awareness, whereas mean values below 2.50 were seen as reflecting unawareness. This benchmark was adopted from the study by Laurence et al. (2012).

Further analysis using Logistic Regression was conducted to find the significant factors influencing the level of awareness of caffeine consumption effects among students. Logistic regression analysis is appropriate for the binary classification, which includes two class values for the dependent variable, aware or not of caffeine consumption among the students. The general form of the logistic regression model is:

$$log\left[\frac{P(x)}{1-P(x)}\right] = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \tag{1}$$

where,
p = probability of success dependent variable
ε = base of natural logarithms
β_o = constant of the equation
$\beta_1,\beta_2,\beta_3,\beta_4$ = the coefficient of independent variable
X_1 = Gender
X_2 = Knowledge
X_3 = Consumption Effect
X_4 = Caffeine Consumption

Thus, this logistic regression analysis is fitted to the empirical data to identify the significant factors influencing the level of awareness of the effects of caffeine consumption on students.

3. Results and Discussion

As mentioned earlier, this study used 397 respondents from a public university student in Negeri Sembilan. The association between categories of daily caffeine drinkers and gender was analysed to capture the respondents' overall characteristics.

Table 1: Association between categories of Daily Caffeine Drinker and Genders

Gender	Drink Caffeinated Daily		Total
	Yes	No	
Male	109	58	167
Female	97	133	230
Total	206	191	397
Chi-square Test, p-value = 0.000			

According to Table 1, 109 male respondents reported drinking caffeine daily, compared to 97 female respondents. However, 133 female respondents claimed they did not drink caffeine daily, compared to 58 male respondents. Generally, it shows that most males consume caffeine daily compared to females. Additionally, the chi-square test of independence yielded a *p*-value less than the chosen significance level of 0.05, indicating sufficient evidence of an association in caffeine intake between genders. These findings align with Dillon et al. (2019) research, which also reported higher caffeine consumption among males than females, but contradict Khan's (2019) findings, which suggested that females consume more caffeine than males. This can be attributed to the biological and physiological differences between genders, with men metabolising caffeine more rapidly than women. As a result, men develop a higher tolerance and tend to consume larger quantities of caffeine (dePaula & Farah, 2019).

3.1. Level of Awareness of Caffeine Consumption Effect

The data in Figure 1 unequivocally proves that approximately 40.3% of respondents firmly believe that reducing their daily caffeine intake will make them happier. Furthermore, a decisive 33.5% of respondents firmly agree that cutting back on caffeine can incontrovertibly help them avoid headaches and anxiety. Additionally, 34.8% of respondents firmly believe that reducing caffeine will improve their health. Moreover, 36.5% of respondents firmly agreed that excessive caffeine consumption will unquestionably harm their health. Furthermore, a decisive 35.8% of respondents strongly agree that they will undeniably inform others about the effects of caffeine. 35.3% of respondents adamantly agree that they are resolutely considering lowering their caffeine intake for their health. Additionally, 39.6% of respondents firmly agree that caffeine can negatively affect their health. Moreover, 43.6% of respondents firmly agree that they fully understand the potential health risks of drinking caffeine. Finally, 44.3% of respondents are divided on whether they continue consuming caffeine daily. From the overall analysis, it can be conclusively stated that almost 50% of respondents agreed with the caffeine consumption effect highlighted.

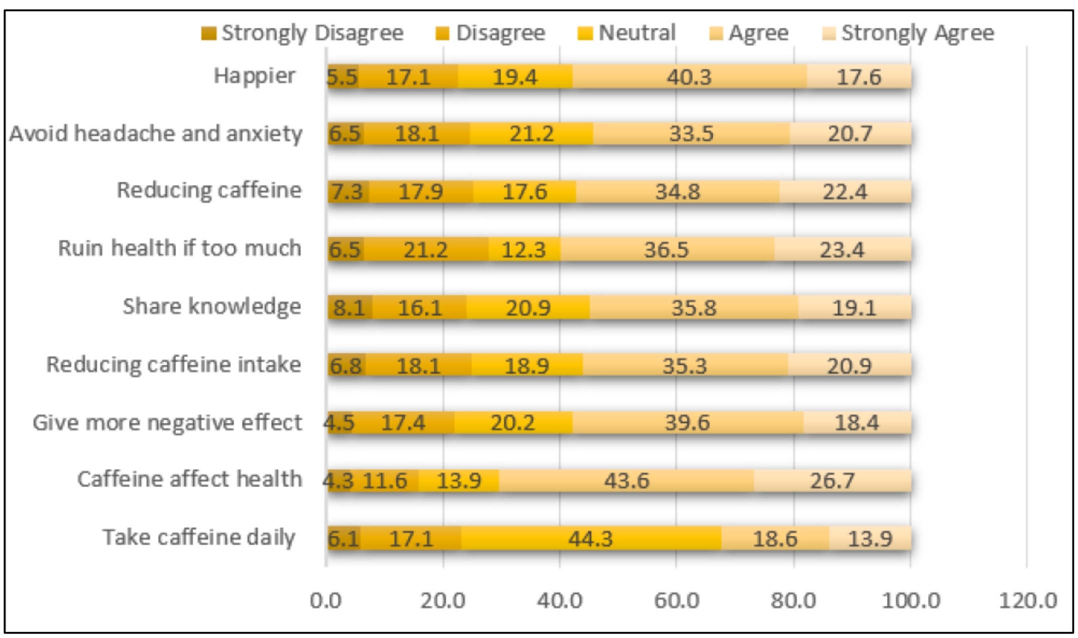


Figure 1: Level of agreement on the caffeine consumption effect

The respondents' awareness of the caffeine consumption effect is demonstrated in Table 2. Out of the 397 respondents involved in the study, 276 (69.5%) reported that they are aware of the effects of caffeine consumption on their health, while 121 (30.5%) of the respondents are not aware. This outcome suggests that individuals become aware of caffeine's side effects after experiencing them firsthand. As noted by Santos et al. (2017), individuals who consume large amounts of caffeine may experience systemic damage such as irregular heart rate, increased ventilation, and anxiety. Hazar (2019) mentioned that people tend to be aware of the side effects of caffeine when they consume it as part of their daily routine.

Table 2: Awareness of caffeine's consumption effect

Level of Awareness	Frequency	%
Aware	276	69.5
Not Aware	121	30.5
Total	397	100.0

However, it is important to note that some respondents who do not consume caffeine regularly may be less aware of its effects. This emphasizes the need for targeted education and awareness campaigns to bridge knowledge gaps and ensure a widespread understanding of caffeine-related health considerations. In this scenario, it is imperative for the responsible party, particularly the student representative, to spearhead an awareness campaign to enhance understanding of caffeine's effects.

3.2. Factor Contributing to Level of Awareness of Caffeine Consumption Effect

3.2.1 Checking model adequacy

There are some assumptions to be satisfied when conducting logistic regression. The dependent variable must be a binary outcome, either aware or unaware of the effects of caffeine consumption. The independent variable can be measured by continuous variables such as knowledge, caffeine consumption, and consumption effect or categorical variables such as gender. Third, the observations should not be derived from many measurements of the same individual or be related to each other in any way. Since, in this study, observations were independent and unrelated to each other, the third assumption is fulfilled. Fourth, there should be no multicollinearity. As shown in Table 3, all variables exhibit tolerance values above 0.01 and VIF values under 10, indicating that multicollinearity is not a concern.

Table 3: Multicollinearity Test

Variable	Tolerance	VIF
Knowledge	0.965	1.036
Consumption Effect	0.883	1.132
Caffeine Consumption	0.902	1.109

Lastly, based on Table 4, the lowest probability value is 0.00192. This result indicates that there are no outliers in the dataset. In conclusion, as all the assumptions were met, the study proceeded to identify the significant variable using logistic regression.

Table 4: Mahalanobis Test

Mahalanobis Value	Probability
14.60134	0.00192
11.97205	0.00251
10.68081	0.00479

3.2.2 Logistic regression analysis and model evaluation

This study has identified several factors, such as gender, knowledge, consumption effect, and caffeine intake, as potential contributors to the level of awareness. However, there are only two statistically significant variables using the logistic regression model: knowledge and consumption effects (Table 5). This is because their *p*-values are 0.005 and 0.003, less than 0.05. Therefore, a significant relationship exists between knowledge and consumption effect and awareness of the effects of caffeine consumption. Knowledge and consumption effects are significant factors that influence the level of awareness of caffeine's consumption effects. The other factors were deemed insignificant and have been removed from the model.

Table 5: Logistic Regression Model

Variable	**Coefficient**	**p-value**	**Wald test**	**Odds Ratio**
Constant	-3.383	0.003	8.755	0.034
Gender (1)	0.108	0.743	0.108	1.114
Knowledge	0.703	0.005	7.717	2.020
Consumption effect	0.805	0.003	8.710	2.236
Caffeine Consumption	-3.383	0.227	1.460	0.723

The odds ratio represents the likelihood that respondents will be aware of the effects of caffeine consumption. Odds reflect the chances that something will or will not happen. Based on the table above, respondents knowledgeable about caffeine's consumption effects are 2.020 times more likely to be aware of these effects than those without knowledge. In addition, respondents who experience the consumption effect are 2.236 times more likely to be aware than respondents who do not experience the consumption effect.

Next, Wald statistics evaluate models based on a best-fit criterion. This approach assesses the statistical significance of each independent variable. The Wald statistics in Table 5 indicate that only the consumption effect and knowledge are significant to the model, as their *p*-value is less than the 0.05 cut-off point.

Additionally, students' knowledge of caffeine and the ingredients of their drinks can help them recognise its drawbacks.

The logit expressions for the full model are as below:

$$log\left[\frac{P(x)}{1-P(x)}\right] = -3.383 + 0.703\ Knowledge + 0.805\ Consumption\ Effect \qquad (2)$$

4. Conclusion and Recommendations

Caffeine, found in coffee, tea, and energy drinks, has benefits like reducing disease risk but can also lead to disrupted sleep, anxiety, and potential cardiovascular issues. It is important to be cautious about consumption. Understanding its effects can help people, especially students, live healthily.

Results have shown that men have been proven to consume more caffeine than women, and the majority of respondents are cognizant of the negative effects of caffeine. Encouragingly, 69.5% of our participants recognised that caffeine could impact their future health. However, it is vital to note that 30.5% of respondents appear to lack awareness of caffeine's side effects. It is encouraging that most respondents are aware of the effects. However, efforts should be made to increase awareness throughout the community about both the benefits and drawbacks of caffeine consumption. Furthermore, the logistic regression model indicates that the consumption effect and knowledge are the significant factors contributing to the level of awareness of the side effects of consuming caffeine. This finding indicates that students are more aware if they are knowledgeable about caffeine.

To truly understand the impact of caffeine consumption, the health ministry must embark on an education campaign. This can involve powerful tools like videos and posters to enlighten people about the harmful effects of excessive caffeine intake. Beginning at a smaller level, such as with a student representative initiating an awareness campaign, can significantly impact. Furthermore, it is important to encourage the community to share knowledge about the effects of caffeine consumption during gatherings and social circles.

Future research could potentially incorporate other factors that are perceived to be significant in relation to people's awareness of the effects of caffeine consumption. Examples of such factors may include environment, lifestyle, study background, and economic considerations, all of which are essential in emphasising and assessing individuals' comprehension of the adverse effects of caffeine. Since this study only identified one significant influencing factor on awareness of caffeine's side effects, it is important to consider these contextual factors. Doing so may lead to a more comprehensive understanding and determination of the most influential factor that can assist in raising awareness within the community.

Acknowledgment

We want to express our heartfelt gratitude to all team members for their unwavering encouragement and dedication, which led to the successful publication of this article.

References

Arul Prakasam, K. C., Glady Gloria Grant, C. J., Aarthy, P., Feba S., Harish, M., & Azad, V. P. (2002). Knowledge, attitude, and practice on caffeine and caffeinated beverages consumption among college students. *International Journal of Pharmacy & Pharmaceutical Research*. 25(3): 329–339.

Aslam, H. M., Mughal, A., Edhi, M. M., Saleem, S., Rao, M. H., Aftab, A., Hanif, M., Ahmed, A., & Khan, A. M. H. (2013). Assessment of pattern for consumption and awareness regarding energy drinks among medical students. *Archives of Public Health*, 71:1–11.

Astorino, T. A., & Roberson, D. W. (2010). Efficacy of acute caffeine ingestion for short-term high-intensity exercise performance: a systematic review. *The Journal of Strength &*

Conditioning Research, 24(1):257–265.

Celi, E., Espinoza, C., Paredes, A., Fabara, J., Velin, D., & Montenegro, M. (2022). The impact of caffeine and energy drinks on health, especially in high-performance athletes. *Health Science Journal*, 16:1–17.

Cochran, W. G. (1977). Sampling techniques. John Wiley & Sons.

dePaula, J., & Farah, A. (2019). Caffeine consumption through coffee: Content in the beverage, metabolism, health benefits and risks. *Beverages*, 5(2):37.

Dillon, P., Kelpin, S., Kendler, K., Thacker, L., Dick, D., & Svikis, D. (2019). Gender differences in any-source caffeine and energy drink use and associated adverse health behaviors. *Journal of Caffeine and Adenosine Research*, 9(1):12–19.

Ginting, S. S. B., Astiarani, Y., Santi, B. T., & Vetinly, V. (2022). Tingkat pengetahuan efek konsumsi kafein dan asupan kafein pada mahasiswa. *Journal of Nutrition College*, 11(4):264–271.

Hazar, A. P. B. (2019). Gambaran pengetahuan dan perilaku pada barista terhadap hipertensi di kedai kopi cileungsi (Unpublished doctoral dissertation). Universitas Kristen Indonesia.

Juliano, L. M., Huntley, E. D., Harrell, P. T., & Westerman, A. T. (2012). Development of the caffeine withdrawal symptom questionnaire: caffeine withdrawal symptoms cluster into 7 factors. *Drug and Alcohol Dependence*, 124(3):229–234.

Khan, N. (2019). Caffeinated beverages and energy drink: Pattern, awareness and health side effects among omani university students. *Biomed Res*, 30(1):113–121.

Laurence, G., Wallman, K., & Guelfi, K. (2012). Effects of caffeine on time trial performance in sedentary men. *Journal of Sports Sciences*, 30(12):35–40.

Pandey, R., Ganatra, V., Jamnik, S., Goel, P., Goyal, P., Xuan, C. L., Kee, D. M. H., Mein, H. H., En, L. X., & Zen, L. J. (2021). An empirical study on customer satisfaction, perception, and brand image in starbucks coffee in india, asia. *International Journal of Tourism and Hospitality in Asia Pasific (IJTHAP)*, 4(1):53–63.

Rodda, S., Booth, N., McKean, J., Chung, A., Park, J. J., & Ware, P. (2020). Mechanisms for the reduction of caffeine consumption: What, how and why. *Drug and Alcohol Dependence*, 212: 108024.

Santos, L. C., Ruiz-Oliveira, J., Silva, P. F., & Luchiari, A. C. (2017). Caffeine dose–response relationship and behavioral screening in zebrafish. *The Question of Caffeine*, 21:87–105.

CLASSIFICATION OF AIR QUALITY IN THE KLANG VALLEY USING K-MEANS CLUSTERING

Nur Aliah Amira Abdul Hamid[1], Siti Khairun Najwa Kamil[2] and Noorezatty Mohd Yusop[3,*]
[1,2,3]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Cawangan Negeri Sembilan Kampus Seremban, 70300 Negeri Sembilan
*noorezatty@uitm.edu.my

Abstract

Air pollution is a major global concern, and its adverse impacts are most likely to affect urban areas. The Klang Valley, located on the southwest coast of Peninsular Malaysia, is also affected by this global issue. As one of the most industrialized and populated areas, controlling air pollution is a challenge. Thus, this paper aims to classify air monitoring stations in the Klang Valley into distinct clusters based on six major air pollutants. Data on pollutant levels collected by the Department of Environment from air monitoring stations throughout the region was utilised. Fourteen stations representing seven areas in Klang Valley were assessed in two weather conditions: the dry season (May-September) and the rainy season (November-March), typically known as monsoon seasons in Malaysia. Two clusters of air monitoring stations were identified using K-Means clustering. The first cluster, comprising four stations, showed better air quality with lower pollutant levels. In contrast, the second cluster, which includes ten stations, showed higher pollutant levels. However, the pollutant levels in both clusters were within the permissible limits according to the Malaysian Ambient Air Quality Guidelines. Furthermore, the location of the monitoring station can influence pollutant levels, whereas seasonal variations (dry or rainy) have a lesser impact. Consistent monitoring is crucial for tracking air pollution changes and adjusting policies accordingly.

Keywords: Air pollution, Clustering Algorithm, K-Means Cluster Analysis

1. Introduction

Pollution issues, particularly air pollution pose major global challenges with urban areas particularly vulnerable to its negative consequences. There are a variety of air pollutants and the release into the air can cause serious health concerns and can even be deadly in small concentrations. In 2021, air pollution caused 8.1 million fatalities worldwide, making it the second most significant cause of death, including children under the age of five (Health Effect Institute, 2024).

Monitoring air quality is essential to protect public health, and ensuring compliance with regulations. It aligns with the World Health Organization's Sustainable Development Goals (SDGs), particularly Goal 3 (Good Health and Well-being) and Goal 11 (Sustainable Cities and Communities), by emphasizing the importance of monitoring air quality to safeguard public health and promote sustainable urban development. Consistent monitoring is crucial for tracking air pollution changes and adjusting policies accordingly.

Currently, the Air Pollution Index (API) is used as an indicator to monitor air quality status. The API value in Table 1 is calculated from six main air pollutants; Particulate Matter of less than 10 Microns (PM_{10}), Particulate Matter of less than 2.5 Microns ($PM_{2.5}$), Ozone (O_3), Carbon Monoxide (CO), Nitrogen Dioxide (NO_2) and Sulphur Dioxide (SO_2).

Table 1: Air pollution index

API Range	Air Pollution Status
0-50	Good
51-100	Moderate
101-200	Unhealthy
201-300	Very Unhealthy
>300	Hazardous

Being one of the most densely populated and industrialized regions in Malaysia, the Klang Valley is increasingly faced with problems relating to air pollution. Klang Valley measures 2911.5 square kilometres in area and is located over several districts in Selangor and the Federal Territory of Kuala Lumpur. Potential sources of air pollution in Klang Valley include biomass burning/smoke (31.6%), soil dust (10.3%), industries (8.4%), motor vehicles (7.1%), sea spray (6.3%) and secondary sulphate (5.5%) (Elias et al., 2023) and the sources continuously contribute to the deterioration of air quality in the Klang Valley. The air monitoring stations measuring pollution levels in the region are located at Batu Muda, Petaling Jaya, Cheras, Shah Alam, Klang, Banting, Kuala Selangor, and Putrajaya.

In the year 2022, the API index for Klang Valley areas mostly fell within the moderate range compared to API readings in the good range as shown in Figure 1. Putrajaya recorded the highest number of good API readings at 98 days, while Klang recorded the highest number of moderate API readings at 360 days and Cheras recorded the highest unhealthy API reading at 5 days with most days falling within moderate status Deparment of Environment (2023). Relying solely on API assessments provides a narrow perspective on pollutant dynamics, overlooking their intricate patterns and interactions. Pollution levels can fluctuate across regions and weather conditions. Classifying monitoring stations based on pollution patterns helps target interventions to address specific pollution sources and patterns effectively.

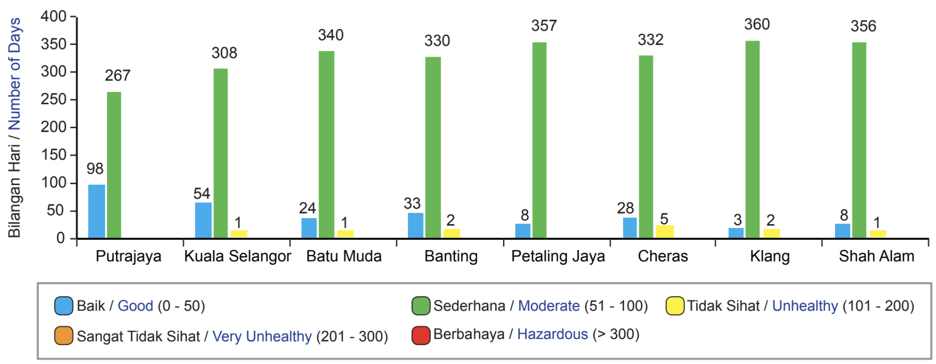


Figure 1: Air quality status in Klang Valley in 2022

Research interest towards air pollution analysis has increased due to its potential to protect public health, preserve environmental sustainability, and ensure compliance with air quality regulations. Previous studies utilizing air pollution data aim to understand pollutant sources (Richard et al., 2023), to predict air quality (Kaur et al., 2023; Vyas et al., 2023) or identify pollutant sites (Goyal et al., 2021). Statistical analysis is essential to simplify the complexities of air pollution data and is used in developing predictive models, identifying factors of air quality and describing pollutant levels. Additionally, advanced methods such as clustering algorithms are employed to reveal hidden patterns.

Clustering algorithms are used to group data points with common attributes, enabling the identification of distinct pollution patterns and sources. This technique assists in categorizing sites with similar pollution profiles, facilitating targeted interventions and policy implementations (Mohamad et al., 2015; Shafi and Waheed, 2020; Sunori et al., 2021). The use of clustering techniques enhances understanding of air quality dynamics. This information, combined with ongoing efforts to minimise air pollution, represents alternate solutions to provide clean air and a sustainable environment for current and future generations. Hence, this paper seeks to employ clustering algorithms to obtain valuable insights into air pollution in urban areas such as Klang Valley enabling effective and targeted environmental interventions.

2. Literature Review

Air pollution is a major environmental issue that has severe health implications. Previous research has been conducted to distinguish the levels of air pollution in different areas, as well as to investigate the cause of high pollutant concentrations. The study conducted by Nguyen and Kim (2006) assessed the regional and temporal distribution patterns of SO_2 in Korea. Data from four different groups of air quality monitoring stations (AQMS) was utilized, encompassing the period spanning from 1998 to 2003. The

AQMS covered densely populated areas in seven major cities (with urban traffic and urban environments), as well as less populated areas in nine major provinces (with suburban environments and rural environments). The result showed significant differences in the average levels of SO_2 among all four groups of stations. A comprehensive analysis comparing pairs of stations (urban traffic vs. urban environment and suburban environment vs. rural environment) revealed that urban traffic stations indicated significantly higher concentrations of SO_2 compared to urban environment stations.

Azmi et al. (2010) on the other hand, assessed air quality patterns in Klang Valley. Data on five major air pollutants (PM_{10}, CO, SO_2, O_3, and NO_2) was gathered from monitoring stations in Petaling Jaya, Shah Alam, and Gombak. These data were compared with benchmark stations in Jerantut, Pahang. Petaling Jaya has higher levels of CO, NO_2, and SO_2 as a consequence of high traffic volume and industrialization. Regional factors such as biomass burning and sunlight affect PM_{10} and O_3 concentration in the area. Although pollutant levels are lower than the permissible limit, they are significantly higher compared to the benchmark station, indicating localized pollution sources.

Furthermore, Anand et al. (2019) examined the levels of particulate matter pollutants ($PM_{2.5}$ and PM_{10}) in four major Indian cities: Delhi, Mumbai, Pune and Ahmedabad, in the North-West region of India. Subsequently, employing data from SAFAR observational network to determine whether geographical location influences particulate matter variability. Besides, the study revealed that although Delhi has the highest concentration of particulate matter, Mumbai shows a higher proportion of $PM_{2.5}$ as a percentage of PM_{10} (60%) than Delhi (50 %). Differences were caused by the fact that Delhi is on land whereas Mumbai is on the coast, which affects the dispersion of air pollution.

In contrast, Mohd Anuar et al. (2021) explored the concentrations of SO_2 in the Universiti Tun Hussein Onn Malaysia (UTHM) Campus Area in Batu Pahat. The study examined air pollution levels in thirteen different areas in different time frames and discovered significant variations in SO_2. The highest recorded levels were detected in the afternoon at TDI Residential College A, with a reading of 0.198ppm, whereas morning measurements, particularly low at University Health Centre with a value of 0.03ppm, indicated changes associated with the distance from pollution sources.

These studies highlight that the variation in air pollution levels across various geographic regions is driven by factors such as burning activities, manufacturing activities, and traffic density.

3. Methodology

3.1. *Source of Data*

This study used secondary data obtained from the Malaysian Department of Environment (DOE). Data on the concentrations of six major air pollutants were obtained during the dry season (May - September) and the rainy season (November - March) in 2022. A total of 2407 records were utilized to generate average concentration values for each pollutant in every monitoring station at the Klang Valley, Selangor, specifically in Cheras, Putrajaya, Shah Alam, Batu Muda, Banting, Klang and Petaling Jaya.

3.2. *Description of Variables*

In the context of assessing air quality, six indicators were used as depicted in Table 2. These indicators provided information about different aspects of air characteristics.

Table 2: Scale of measurements for variables

Variable	Unit	Measurement
Air Monitoring Station	-	Nominal
Particulate Matter ($PM_{2.5}$)	μg/m^3	Ratio
Particulate Matter (PM_{10})	μg/m^3	Ratio
Nitrogen Dioxide (NO_2)	ppm	Ratio
Sulphur Dioxide (SO_2)	ppm	Ratio
Carbon Dioxide (CO)	ppm	Ratio
Ozone (O_3)	ppm	Ratio

Table 3 contains information on the cluster members, which are the air monitoring stations in Klang Valley. Fourteen monitoring stations were assessed, with data collected from selected areas in Klang Valley during two seasons: dry (May-September) and rainy (November-March). This setup allows for data comparisons between the two time periods.

Table 3: Air monitoring stations in Klang Valley

Station ID	Station Name	Season
BM1	Batu Muda	Dry
BM2	Batu Muda	Rainy
C1	Cheras	Dry
C2	Cheras	Rainy
PW1	Putrajaya	Dry
PW2	Putrajaya	Rainy
PJ1	Petaling Jaya	Dry
PJ2	Petaling Jaya	Rainy
SA1	Shah Alam	Dry
SA2	Shah Alam	Rainy
K1	Klang	Dry
K2	Klang	Rainy
B1	Banting	Dry
B2	Banting	Rainy

3.3. *K-Means Clustering Algorithm*

The K-Means clustering algorithm is employed to categorise monitoring stations into different groups based on similar properties. K-Means algorithm operates by taking an input of objects (in this study the monitoring stations), denoted as C and an integer value K, ultimately producing a partition of C into subsets $C_1, C_2, \ldots, C_k$. The approach is regarded as an optimization criterion aimed at minimizing the sum of squared distances of objects within their respective clusters from the cluster centres (Chen et al., 2002). The sum of squared criterion is defined by the cost function,

$$W(C_i) = \sum_{r=1}^{|C_i|} \sum_{s=1}^{|C_i|} (d(x_r^i, x_s^i))^2 \tag{1}$$

Where
x_r^i = The r^{th} element of C_i
x_s^i = The s^{th} element of C_i
$|C_i|$ = Number of elements in C_i
$d(x_r^i, x_s^i)$ = Distance between a data point x_r^i and x_s^i

The K-Means algorithm calculates the centroid of each cluster C_i (denoted by x_i) and optimizes the cost function that can be defined as;

$$W(C_i) = \sum_{r=1}^{|C_i|} (d(x^i x_r^i))^2 \tag{2}$$

The K-Means algorithm consists of several steps:

(1) Determine the K number of clusters in which objects are categorized, assigning them to one of these K groups.
(2) Allocate each object to the group that has the closest centroid (mean) with respect to Euclidean distance. The Euclidean distance can be defined as the distance between x_i and y_i in n-dimensional space,

$$d_E(x_i, y_i) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3}$$

Here, n is the number of indicators.

(3) Once all the objects have been assigned, recalculate the centroid for the clustering gaining the new object and adjust the centroid for the cluster losing the object.

(4) Repeat steps 2 and 3 until some of the convergence is met. That is when centroids reach a state where they remain static. The convergence criteria either no or minimal reassignment of data points to new cluster centres and a minimal decrease in the sum of squared error.

3.4. *Cluster Validity Index*

Finding the optimal number of clusters in a dataset is a critical aspect of clustering analysis. Cluster validity indices play a significant role in this process by providing objective criteria to evaluate and compare clustering results for different numbers of clusters.

Krzanowski-Lai Index assists in finding the optimal number of clusters in the data Syahputri et al. (2024). This method is defined by the difference between cluster k with $k-1$ using the equation in (4).

$$DIFF(k) = (k-1)^{\frac{2}{p}} W_{k-1} - k^{\frac{2}{p}} W_k \tag{4}$$

The cluster number, k is selected by maximizing the quantity based on equation (5).

$$KL(k) = |\frac{DIFF(k)}{DIFF(k+1)}| \tag{5}$$

Where:
p = Number of dimensions (indicator)
k = Number of clusters

4. Result and Discussion

In this section, the clusters of air monitoring stations are identified. There are fourteen air monitoring stations in total, spanning seven areas across two seasons. The optimal number of clusters was determined using the Krzanowski-Lai Index, and K-Means analysis was employed to identify the cluster members. Additionally, a 2D clustering plot is used to visually represent the distinct clusters.

4.1. *Determining the Optimal Number of Cluster*

Krzanowski-Lai Index was utilized for this purpose. Table 4 shows the Krzanowski-Lai Index score for cluster number 2 to 5. The highest score 1.8513, was obtained at cluster number 2, affirming that the optimal number of clusters is 2.

Table 4: Krzanowski-Lai index score

Cluster Number	Score
2	**1.8513**
3	1.4078
4	0.4893
5	0.9142

Figure 2 displays the score in a plot with the x-axis denoting the cluster count and the y-axis indicating the corresponding Krzanowski-Lai Index score. The plot demonstrates that the highest point occurs at 2.

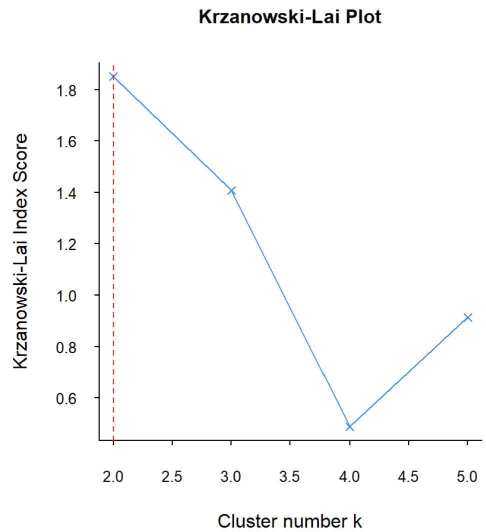


Figure 2: Krzanowski-Lai index plot

4.2. *Identifying Monitoring Stations by Cluster*

The 2D clustering plot in Figure 3 visually presents the outcome of the K-Means analysis. By employing K-Means analysis, two unique clusters of air monitoring stations were identified. Each point on the plot corresponds to a monitoring station at a specific location, with the numbers 1 or 2 representing seasonal points, dry or rainy season. Points of the same color are grouped to form a cluster, positioned close to their cluster centroid. The cluster numbers (e.g., 1, 2, etc.) are assigned according to the sequence in which the centroids are initialized and subsequently updated during the clustering algorithm's iterations. Consequently, these numbers are somewhat arbitrary and do not inherently correspond to any specific criteria

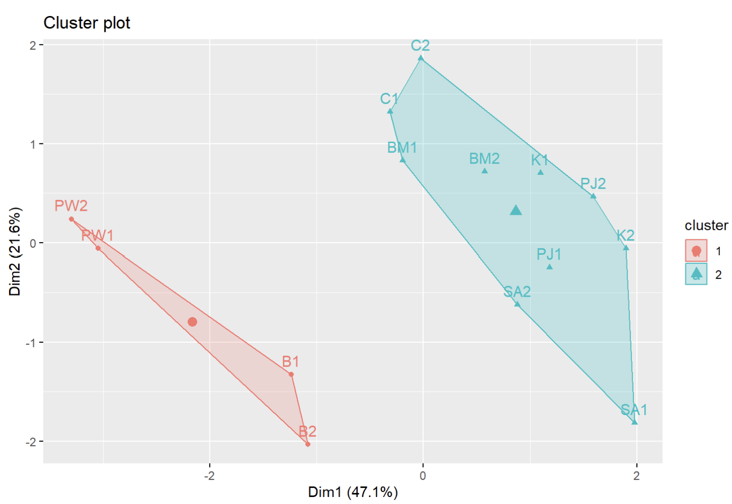


Figure 3: Clusters of air monitoring stations in Klang Valley

The results of the clustering analysis are summarized in Table 5, which illustrates the distribution of stations across different seasons for each obtained cluster. The monitoring stations are divided into two clusters. Cluster 1 contains four stations: Putrajaya and Banting, during both the dry and rainy seasons. The remaining ten stations fall into Cluster 2, which includes Batu Muda, Cheras, Petaling Jaya, Shah Alam, and Klang, also encompassing both the dry and rainy seasons.

Table 5: Cluster members

Cluster	Station ID	Station Name	Season
1	PW1	Putrajaya	Dry
	PW2	Putrajaya	Rainy
	B1	Banting	Dry
	B2	Banting	Rainy
2	BM1	Batu Muda	Dry
	BM2	Batu Muda	Rainy
	C1	Cheras	Dry
	C2	Cheras	Rainy
	PJ1	Petaling Jaya	Dry
	PJ2	Petaling Jaya	Rainy
	SA1	Shah Alam	Dry
	SA2	Shah Alam	Rainy
	K1	Klang	Dry
	K2	Klang	Rainy

Table 6 presents summary statistics for six air pollutants across Clusters 1 and 2, focusing on average concentrations. The table indicates that air pollutant levels in both clusters are within permissible limits. Cluster 2 shows higher average concentrations for four pollutants (PM_{10}, $PM_{2.5}$, CO, and NO_2), whereas Cluster 1 exhibits higher pollutant levels for two types of pollutants (O_3 and SO_2).

Table 6: Average pollutant levels across two clusters

Pollutants	Average exposure limit	Cluster	
		1	2
Particulate Matter ($PM_{2.5}$)	$260\mu g/m^3$	24.0097	**28.5323**
Particulate Matter (PM_{10})	$260\mu g/m^3$	15.8125	**19.3600**
Sulfur Dioxide (SO_2)	$0.04ppm$	**0.0015**	0.0013
Nitrogen Dioxide (NO_2)	$0.04ppm$	0.0088	**0.0135**
Ozone (O_3)	$0.06ppm$	**0.0197**	0.0154
Carbon Monoxide (CO)	$9ppm$	0.5076	**0.7015**

Statistics in Table 6 indicate that the air monitoring stations in Cluster 1, which exhibit lower concentration values, are in Putrajaya and Banting for both the dry and rainy seasons. It shows that changes in the seasons do not affect the pollutant levels. Although it is anticipated that rainfall could lower the pollutant levels, however shorter exposure to rainfall which is less than an hour, keeps some pollutant level to remain constant (Hilario et al., 2022). Generally, Klang Valley develop short-duration rainfall between 15 to 45 minutes (Mamun et al., 2018).

Putrajaya is primarily designed as an administrative capital with extensive parks, green spaces, and water features. It aims to become a "green city" by implementing eco-friendly structures, promoting public transportation, and using renewable energy sources, all contributing to pollution reduction (Abd. Razak, 2014). In contrast, Banting, a suburban village nestled near hills, forests, and farmland, hosts industrial zones that contribute to air pollution, particularly during El Niño periods. However, it achieved a good Air Pollution Index (API) status in 2022, possibly due to pollutant dispersion in agricultural and forested areas (Shaadan et al., 2018).

In Klang Valley, stations like Cheras, Petaling Jaya, Batu Muda, Shah Alam, and Klang form a different cluster, showing higher pollutant concentrations and poorer air quality attributed to urban growth, industrial activities, and heavy traffic. Cheras had the highest unhealthy API, with significant carbon monoxide emissions from motor vehicles, while Klang reported the highest moderate API in 2022 (Deparment of Environment, 2023). Despite seasonal haze from fires in Sumatra and Borneo affecting the region, pollutant concentrations in Klang Valley generally remain within acceptable limits due to dispersion and absorption in natural environments (Deparment of Environment, 2023).

5. Conclusion

In conclusion, the analysis of air quality across different areas within the Klang Valley reveals diverse environmental dynamics and their impact on pollutant concentrations. The monitoring stations were successfully divided into two distinct clusters, with each member within a cluster exhibiting similar air pollution characteristics. The first cluster includes stations in Putrajaya and Banting, which are recognized for their comparatively better air quality and lower average pollution levels. In contrast, the second cluster comprises stations in more urbanized areas like Cheras, Petaling Jaya, Batu Muda, Shah Alam, and Klang, where higher average pollution levels are observed due to industrial emissions and vehicular pollution. Based on the pollutant levels, Cluster 1 has higher levels of SO_2 and O_3 which is indicating to higher concentrations of certain gases. Encouraging people to walk, bicycle, or use public transport can help reduce car emissions. On the other hand, Cluster 2 has higher level of PM_{10}, $PM_{2.5}$, NO_2, and CO. Strict regulations can help control these pollutants. Construction companies, for example, can take steps to reduce pollution by installing dust control measures on building sites and upgrading industrial equipment. Moving forward, it is essential to strengthen urban planning strategies and implement stricter regulatory measures to manage air quality effectively across Klang Valley. Continuously monitoring air quality across all clusters and enhancing public awareness about the impact of pollution on health and the environment are crucial steps in achieving sustainable development and improving quality of life in the region.

Acknowledgment

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

Abd. Razak, M. R. (2014). Putrajaya green cities: constraints and mitigation strategies towards low carbon cities [unpublished thesis]. Universiti Teknologi Malaysia.

Anand, V., Korhale, N., Rathod, A., and Beig, G. (2019). On processes controlling fine particulate matters in four indian megacities. *Environmental Pollution*, 254:113026.

Azmi, S. Z., Latif, M. T., Ismail, A. S., Juneng, L., and Jemain, A. A. (2010). Trend and status of air quality at three different monitoring stations in the klang valley, malaysia. *Air Quality, Atmosphere & Health*, 3:53–64.

Chen, G., Jaradat, S., Banerjee, N., Tanaka, T., Ko, M. S., and Zhang, M. Q. (2002). Evaluation and comparison of clustering algorithms in analyzing ES cell gene expression data. *Statistica Sinica*, 12(1):241–262.

Deparment of Environment (2023). *Laporan Kualiti Alam Sekeliling (EQR) 2022*. Ministry Of Natural Resources And Environmental Sustainability, Petaling Jaya.

Elias, M. S., Hashim, A., Bahrudin, N. F. D., Sapiee, N. A., Paulus, W., Azman, M. A., Raduian, N. J., and Abdullah, I. M. (2023). Assessment and sources identification of air quality pollution in Klang Valley, Kuala Lumpur, Malaysia. *IOP Conference Series Materials Science and Engineering*, 1285(1):012017.

Goyal, P., Gulia, S., and Goyal, S. (2021). Identification of air pollution hotspots in urban areas-an innovative approach using monitored concentrations data. *Science of The Total Environment*, 798:149143.

Health Effect Institute (2024). *State of Global Air 2024*. Health Effects Institute, Boston, MA.

Hilario, M. R. A., Bañaga, P. A., Betito, G., Braun, R. A., Cambaliza, M. O., Cruz, M. T., Lorenzo, G. R., MacDonald, A. B., Pabroa, P. C., Simpas, J. B., et al. (2022). Stubborn aerosol: why particulate mass concentrations do not drop during the wet season in metro manila, philippines. *Environmental Science: Atmospheres*, 2(6):1428–1437.

Kaur, J., Parmar, K. S., and Singh, S. (2023). Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environmental Science and Pollution Research*, 30(8):19617–19641.

Mamun, A. A., Salleh, M. N. B., and Noor, H. M. (2018). Estimation of short-duration rainfall intensity from daily rainfall values in Klang Valley, Malaysia. *Applied Water Science*, 8(7).

Mohamad, N. D., Ash'aari, Z. H., and Othman, M. (2015). Preliminary assessment of air pollutant sources identification at selected monitoring stations in Klang Valley, Malaysia. *Procedia Environmental Sciences*, 30:121–126.

Mohd Anuar, F. N., Othman, N., and Hamood Altowayti, W. A. (2021). Assessments of concentrations of sulfur dioxide (so2) in universiti tun hussein onn malaysia (uthm) campus area, batu pahat. In *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, pages 1–6.

Nguyen, H. T. and Kim, K. (2006). Evaluation of SO2 pollution levels between four different types of air quality monitoring stations. *Atmospheric Environment*, 40(36):7066–7081.

Richard, G., Izah, S. C., and Ibrahim, M. (2023). Air pollution in the niger delta region of nigeria: sources, health effects, and strategies for mitigation. *Journal of Environmental Studies*, 29(1):1–15.

Shaadan, N., Nazeri, L., Jalani, M., Rahman, N., and Roslan, R. (2018). Data visualization of temporal ozone pollution between urban and sub-urban locations in selangor malaysia. *Journal of Fundamental and Applied Sciences*, 9:490.

Shafi, J. and Waheed, A. (2020). K-means clustering analysing abrupt changes in air quality. *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 3:26–30.

Sunori, S. K., Negi, P. B., Maurya, S., Juneja, P., Rana, A., and Bhawana (2021). K-means clustering of ambient air quality data of uttarakhand, india during lockdown period of covid-19 pandemic. In *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, pages 1254–1259.

Syahputri, Z., Siregar, M. A. P., et al. (2024). Determining the optimal number of k-means clusters using the calinski harabasz index and krzanowski and lai index methods for groupsing flood prone areas in north sumatra. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 9(1):571–580.

Vyas, V., Mehta, K., and Sharma, R. (2023). The nexus between toxic-air pollution, health expenditure, and economic growth: An empirical study using ardl. *International Review of Economics & Finance*, 84:154–166.

EXTREMAL PROPERTIES OF CERTAIN CLASS OF TILTED UNIVALENT ANALYTIC FUNCTIONS

Noor Jannah Nazri[1], Nur Adibah Abdul Razak[2], Sharifah Syazwani Syed Abdul Halim[3], and Abdullah Yahya[4,*]

[1,2,3,4]College of Computing, Informatics and Mathematics, UiTM Cawangan Negeri Sembilan, Kampus Seremban, Malaysia.

*abdullahyahya@uitm.edu.my

Abstract

This research concerns some extremal properties of certain class of univalent analytic functions that include representation theorem and coefficient bound. Let S denote the subclass of univalent functions f in an open unit disc, $E=\{m:|m|<1\}$, given by $f(m)=m+\sum_{n=2}^{\infty}a_n m^n$. The study focuses on the generalised class of tilted univalent analytic functions, $S^*(\delta,t)$ which denoted as $\operatorname{Re}\left\{e^{i\alpha}\frac{mf'(m)}{g'(m)}\right\}>\delta, m\in E,$ where $|\alpha|\le\pi,$ $\cos\alpha>\delta,$ $0\le\delta\le1,$ $-1<t\le1$ and $g'(m)=\frac{m}{(1+tm)(1-m)}$. In this study, the new generalized class of tilted univalent analytic functions, representation theorem and the coefficient bound for the $S^*(\delta,t)$ are obtained by using Herglotz Representation Theorem.

Keywords: Extremal properties, Representation theorem, Coefficient bound

1. Introduction

Complex analysis is a part of mathematical fields that deals with analytic functions of a complex variable. The Geometric Function Theory (GFT) is subdividing of complex analysis which included geometric properties of univalent analytic function, initiated by a German Mathematicians, Ludwig Bieberbach in 1916. As stated by Goodman (1983), an analytic function $f(m)$ is a one-to-one mapping of one region to another in a complex plane.

Let $\mathbb{C}$ be an element of complex number and let $f(m)$ be a complex-valued function of the complex variable m. A function f is called univalent on a domain, $D\in\mathbb{C}$ if the function f is injective, which is for all $m_1,m_2\in D, f(m)=f(m_2)$ implies $m_1=m_2$ (Duren, 1983). Then, Darus (2002) stated that the codomain and range of one-to-one function are in real axis R which need on one axis only. In addition, at a point $m\in D$ is said to be analytic if it is differentiable at every point of some open neighborhood of m_0.

Let A denoted the class of function of the form

$$f(m) = m + a_2 m^2 + a_3 m^3 + a_4 m^4 + \ldots + a_n m^n = m + \sum_{n=2}^{\infty} a_n m^n, \tag{1.1}$$

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

Based on Kaharudin (2011), if $f \in S$ is given by (1.1) and $f \in G_K(\alpha, \delta)$, then

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

where the functions in this class satisfy the condition

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

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

In addition, Yahya, Soh, and Mohamad (2013) stated that if $f \in S$ and $f \in G(\alpha, \delta)$ then

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

where the functions in this class satisfy the condition

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

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

In the present paper, we focused on the generalised class of tilted univalent analytic functions, $S^*(\delta, t)$ which denoted as S and satisfied the condition

$$\operatorname{Re}\left\{ e^{i\alpha} \frac{mf'(m)}{g'(m)} \right\} > \delta, m \in E, \tag{1.2}$$

where $|\alpha| < \pi$, $\cos\alpha > \delta$, $0 \le \delta < 1$, $-1 < t \le 1$, and $g'(m) = \dfrac{m}{(1+tm)(1-m)}$. The main objectives of this paper are to define the new generalized class of tilted univalent analytic functions, to find representation theorem for this class of functions and to determine the coefficient bound for the $S^*(\delta,t)$ by using Herglotz Representation Theorem.

2. Preliminaries

To derive the main result, we apply Herglotz Representation Theorem to obtain Representation Theorem for the $S^*(\delta,t)$.

Theorem 2.1
Let $f \in S$ and $f \in S^*(\delta, t)$, then

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

Proof.
Based on (1.2), let $f'(m)(1+tm)(1-m)$ be written as

$$f'(m)(1+tm)(1-m)=1+\sum_{n=1}^{\infty}p_n m^n. \tag{2.1}$$

Then, (2.1) can be written as

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

By applying into the relation P, we have,

$$e^{i\alpha}\ f'(m)(1+tm)(1-m)-\delta=e^{i\alpha}\left[1+\sum_{n=1}^{\infty}p_n m^n\right]-\delta \tag{2.2}$$

Rearranging (2.2), we have,

$$e^{i\alpha}\ f'(m)(1+tm)(1-m)-\delta-i\sin\alpha=\cos\alpha-\delta+\sum_{n=1}^{\infty}\left(e^{i\alpha}p_n\right)m^n.$$

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

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

Therefore,

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

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

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

We relate the function in P with

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

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

Now, we shall prove our main result.

3. Main Result

Now, we shall focus on the coefficient bound of $S^*(\delta,t)$.

Theorem 3.1
If $f\in S$ and $f\in S^*(\delta,t)$, then

$$|a_n|\le\begin{cases}\dfrac{1}{n}\left(\dfrac{1-t^2+2A_{\alpha\delta}\left[(-1+t)+n(t+1)\right]}{(t+1)^2}\right), & n=2,4,6,\ldots\\[2ex] \dfrac{1}{n}\left(\dfrac{2A_{\alpha\delta}(n-1)+t+1}{(t+1)}\right), & n=3,5,7,\ldots\end{cases}$$

Proof.
Suppose that

$$p\in P\Leftrightarrow p(m)=\int_x\frac{1+xm}{1-xm}\,d\mu(x)\,,\quad |x|=1\,,$$

for some probability measure μ on the unit circle X. Rearranging (2.3) to make $f'(m)$ as the subject,

$$e^{i\alpha}\frac{mf'(m)}{g'(m)}-\delta-i\sin\alpha=p(m)(\cos\alpha-\delta).$$

Then,

$$e^{i\alpha}f'(m)=\frac{g'(m)\left[p(m)(\cos\alpha-\delta)+\delta+i\sin\alpha\right]}{m}\,,$$

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

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

Therefore,

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

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

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

Then, follows that

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

Then,

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

$$= \int_0^m \left[\int_x \frac{1}{(1+xt\varphi)(1-x\varphi)}\left(\frac{1+x\varphi\left[e^{-i\alpha}(\cos\alpha - i\sin\alpha - 2\delta)\right] d\mu(x)}{1-x\varphi}\right)\right] d\varphi,$$

and

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

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

Let $e^{-i2\alpha} - 2\delta e^{-ia} = m,$

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

Rearranging the equation (3.2),

$$f(m) = \int_X \left[\int_0^m \frac{-m + x\varphi m + 1 + m}{(1-x\varphi)^2(1+xt\varphi)}\, d\varphi\right] d\mu(x)$$

$$= \int_X \int_0^m \left[\frac{-m}{(1-x\varphi)(1+xt\varphi)} + \frac{1+m}{(1-x\varphi)^2(1+xt\varphi)}\right] d\varphi\, d\mu(x),$$

Therefore,

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

Next, separate the variable

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

By using partial fraction, we get

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

by replacing,

$$1+e^{-2i\alpha}-2\delta e^{-i\alpha}=2A_{\alpha\delta}\left(\cos\alpha-i\sin\alpha\right)$$

and rearrange the equation,

$$f(m)=\int_X\left[\int_0^m\left[\left(\frac{-e^{-i2\alpha}+2\delta e^{-i\alpha}}{(t+1)}\right)\left(\frac{1}{(1-x\phi)}+\frac{t}{(1+xt\phi)}\right)\right.\right.$$
$$\left.\left.+\left(\frac{2A_{\alpha\delta}e^{-i\alpha}}{(t+1)^2}\right)\left(\frac{t}{(1-x\phi)}+\frac{1+t}{(1-x\phi)^2}+\frac{t^2}{(1+xt\phi)}\right)\right]d\phi\right]d\mu(x). \quad (3.3)$$

From (3.3), we have

$$f(m)=\frac{1}{t+1}\int_0^m\left[\int_X\left[\left(-e^{-i2\alpha}+2\delta e^{-i\alpha}\right)\left(\frac{1}{(1-x\phi)}+\frac{t}{(1+xt\phi)}\right)\right.\right.$$
$$\left.\left.+\left(\frac{2A_{\alpha\delta}e^{-i\alpha}}{t+1}\right)\left(\frac{t-tx\phi}{(1-x\phi)^2}+\frac{1+t}{(1-x\phi)^2}+\frac{t^2}{(1+xt\phi)}\right)\right]d\mu(x)\right]d\phi,$$

$$=\frac{1}{t+1}\int_0^m\left[\int_X\left[\left(-e^{-i2\alpha}+2\delta e^{-i\alpha}\right)\left(\frac{1}{(1-x\phi)}+\frac{t}{(1+xt\phi)}\right)\right.\right.$$
$$\left.\left.+\left(\frac{2A_{\alpha\delta}e^{-i\alpha}}{t+1}\right)\left(\frac{-tx\phi}{(1-x\phi)^2}+\frac{1+2t}{(1-x\phi)^2}+\frac{t^2}{(1+xt\phi)}\right)\right]d\mu(x)\right]d\phi,$$

and

$$\begin{aligned} f(m) = \frac{1}{t+1}\int_0^m \Big[& -\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)\int_X \sum_{n=0}^{\infty} (x)^n \, d\mu(x)(\phi)^n \\ & + \left(-te^{-i2\alpha} + 2t\delta e^{-i\alpha} + \frac{2t^2 A_{\alpha\delta} e^{-i\alpha}}{(t+1)}\right)\int_X \sum_{n=0}^{\infty} (-1)^n (t)^n (x)^n d\mu(x)(\phi)^n \\ & - \frac{2tA_{\alpha\delta} e^{-i\alpha}}{(t+1)} \int_X \sum_{n=0}^{\infty} (n)(x)^n \, d\mu(x)(\phi)^n \\ & + \frac{2A_{\alpha\delta} e^{-i\alpha}(1+2t)}{(t+1)} \int_X \sum_{n=0}^{\infty} (n+1)(x)^n d\mu(x)(\phi)^n \Big] d\phi. \end{aligned} \tag{3.4}$$

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

$$f(m) = \frac{1}{(t+1)^2}\left(\left[\left(-e^{-i2\alpha} + 2\delta e^{-i\alpha} - te^{-i2\alpha} + 2t\delta e^{-i\alpha}\right)(t+1)\right] + 2t^2 A_{\alpha\delta} e^{-i\alpha} + 2A_{\alpha\delta} e^{-i\alpha}(1+2t)\right),$$

and

$$f(m) = \frac{1}{(t+1)^2}\left(\left(-1-2t-t^2\right)e^{-i2\alpha} + \left(2t^2+4t+2\right)\delta e^{-i\alpha} + \left(2t^2+4t+2\right)A_{\alpha\delta} e^{-i\alpha}\right).$$

Substitute $A_{\alpha\delta} = \cos\alpha - \delta$ and $\cos\alpha = \dfrac{e^{i\alpha} + e^{-i\alpha}}{2}$, thus

$$\begin{aligned} f(m) &= \frac{1}{(t+1)^2}\left(\left(-1-2t-t^2+t^2+2t+1\right)e^{-i2\alpha} + \left(2t^2+4t+2-2t^2-4t-2\right)\delta e^{-i\alpha} + \left(t^2+2t+1\right)\right) \\ &= 1 \end{aligned}$$

Then, rewrite the equation (3.4) will yield to

$$\begin{aligned} f(m) = \int_0^m \Bigg[1 & - \left(\frac{e^{-i2\alpha} - 2\delta e^{-i\alpha}}{t+1}\right)\int_X \sum_{n=1}^{\infty} (x)^n \, d\mu(x)(\phi)^n \\ & - \left(\frac{te^{-i2\alpha}(t+1) - 2t\delta e^{-i\alpha}(t+1) - 2t^2 A_{\alpha\delta} e^{-i\alpha}}{(t+1)^2}\right)\int_X \sum_{n=1}^{\infty} (-t)^n (x)^n \, d\mu(x)(\phi)^n \\ & - \frac{2tA_{\alpha\delta} e^{-i\alpha}}{(t+1)^2} \int_X \sum_{n=1}^{\infty} (n)(x)^n \, d\mu(x)(\phi)^n \\ & + \frac{2A_{\alpha\delta} e^{-i\alpha}(1+2t)}{(t+1)^2} \int_X \sum_{n=1}^{\infty} (n+1)(x)^n \, d\mu(x)(\phi)^n \Bigg] d\phi. \end{aligned}$$

Integrating with respect to ϕ gives us,

$$
\begin{aligned}
f(m) = m &- \left(\frac{e^{-i2\alpha} - 2\delta e^{-i\alpha}}{t+1}\right)\int_X \sum_{n=2}^{\infty} (x)^{n-1}\, d\mu(x)\left(\frac{m^n}{n}\right) \\
&- \left(\frac{te^{-i2\alpha}(t+1) - 2t\delta e^{-i\alpha}(t+1) - 2t^2 A_{\alpha\delta} e^{-i\alpha}}{(t+1)^2}\right)\int_X \sum_{n=2}^{\infty} (-t)^{n-1}(x)^{n-1}\, d\mu(x)\left(\frac{m^n}{n}\right) \\
&- \frac{2tA_{\alpha\delta}e^{-i\alpha}}{(t+1)^2}\int_X \sum_{n=2}^{\infty} (n-1)(x)^{n-1}\, d\mu(x)\left(\frac{m^n}{n}\right) \\
&+ \frac{2A_{\alpha\delta}e^{-i\alpha}(1+2t)}{(t+1)^2}\int_X \sum_{n=2}^{\infty} (n)(x)^{n-1}\, d\mu(x)\left(\frac{m^n}{n}\right)\Bigg].
\end{aligned}
$$

Rearranging the equation, we have

$$
\begin{aligned}
f(m) = m + \sum_{n=2}^{\infty}\Bigg(&-\frac{\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)}{(t+1)n}\left[\int_X (x)^{n-1} d\mu(x)\right] \\
&- \left(\frac{te^{-i2\alpha}(t+1) - 2t\delta e^{-i\alpha}(t+1) - 2t^2 A_{\alpha\delta} e^{-i\alpha}}{(t+1)^2 n}\right)\left[\int_X (-t)^{n-1}(x)^{n-1}\, d\mu(x)\right] \\
&- \frac{2tA_{\alpha\delta}e^{-i\alpha}}{(t+1)^2 n}\left[\int_X (n-1)(x)^{n-1}\, d\mu(x)\right] \\
&+ \frac{2A_{\alpha\delta}e^{-i\alpha}(1+2t)}{(t+1)^2 n}\left[\int_X (n)(x)^{n-1}\, d\mu(x)\right]\Bigg)m^n,
\end{aligned}
\tag{3.5}
$$

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

$$
\begin{aligned}
a_n = \Bigg(&-\frac{\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right)}{(t+1)n}\left[\int_X (x)^{n-1} d\mu(x)\right] \\
&- \left(\frac{te^{-i2\alpha}(t+1) - 2t\delta e^{-i\alpha}(t+1) - 2t^2 A_{\alpha\delta} e^{-i\alpha}}{(t+1)^2 n}\right)\left[\int_X (-t)^{n-1}(x)^{n-1}\, d\mu(x)\right] \\
&- \frac{2tA_{\alpha\delta}e^{-i\alpha}}{(t+1)^2 n}\left[\int_X (n-1)(x)^{n-1}\, d\mu(x)\right] \\
&+ \frac{2A_{\alpha\delta}e^{-i\alpha}(1+2t)}{(t+1)^2 n}\left[\int_X (n)(x)^{n-1}\, d\mu(x)\right]\Bigg)m^n.
\end{aligned}
$$

Upon simplification, we have

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

Rearranging the equation,

$$|a_n| = \frac{1}{(t+1)n}\left|\left(-e^{-i2\alpha} + 2\delta e^{-i\alpha} - \left(te^{-i2\alpha} - 2t\delta e^{-i\alpha} - \frac{2t^2 A_{\alpha\delta}e^{-i\alpha}}{(t+1)}\right)(-t)^{n-1}\right.\right.$$
$$\left.\left. - \frac{2ntA_{\alpha\delta}e^{-i\alpha}}{(t+1)} + \frac{2tA_{\alpha\delta}e^{-i\alpha}}{(t+1)} + \frac{2nA_{\alpha\delta}e^{-i\alpha}}{(t+1)} + \frac{4ntA_{\alpha\delta}e^{-i\alpha}}{(t+1)}\right)\right|\left|\int_X (x)^{n-1} d\mu(x)\right|,$$

we have,

$$|a_n| = \frac{1}{(t+1)n}\left|\left(-e^{-i2\alpha} + 2\delta e^{-i\alpha} - \left(-e^{-i2\alpha} + 2\delta e^{-i\alpha} + \frac{2tA_{\alpha\delta}e^{-i\alpha}}{(t+1)}\right)(-t)^{n}\right.\right.$$
$$\left.\left. + \frac{2ntA_{\alpha\delta}e^{-i\alpha}}{(t+1)} + \frac{2tA_{\alpha\delta}e^{-i\alpha}}{(t+1)} + \frac{2nA_{\alpha\delta}e^{-i\alpha}}{(t+1)}\right)\right|\left|\int_X (x)^{n-1} d\mu(x)\right|. \qquad (3.6)$$

Based on (3.6), we will obtain two equations of a_n. One of the equations of a_n when n is an even number starting with $n = 2,4,6,\ldots$

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

and

$$|a_n| = \frac{1}{(t+1)n}\left|\frac{(-1+t)\left[(t+1)\left(e^{-i2\alpha} - 2\delta e^{-i\alpha}\right) - 2tA_{\alpha\delta}e^{-i\alpha}\right]}{(t+1)} + 2nA_{\alpha\delta}e^{-i\alpha}\right|\left|\int_X (x)^{n-1} d\mu(x)\right|,$$

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

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

and

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

thus

$$\begin{aligned}|a_n| &= \frac{1}{(t+1)n}\left|\frac{1-t^2+2A_{\alpha\delta}e^{-i\alpha}\left[(-1+t)+n(t+1)\right]}{(t+1)}\right|\int_X |(x)^{n-1}|\, d\mu(x)\\ &\le \frac{1}{(t+1)n}\left(\frac{1-t^2+2A_{\alpha\delta}\left[(-1+t)+n(t+1)\right]}{(t+1)}\right)\int_X |(x)^{n-1}|\, d\mu(x),\end{aligned}$$

and

$$|a_n| = \frac{1}{n}\left(\frac{1-t^2+2A_{\alpha\delta}[(-1+t)+n(t+1)]}{(t+1)^2}\right), \quad n = 2,4,6,\ldots$$

Based on (3.6), another equation of a_n when n is an odd number starting with $n = 3,5,7,\ldots$

$$\begin{aligned}|a_n| = \frac{1}{(t+1)n}\Bigg|\Bigg(&-e^{-i2\alpha}+2\delta e^{-i\alpha}-\left(te^{-i2\alpha}-2t\delta e^{-i\alpha}-\frac{2t^2A_{\alpha\delta}e^{-i\alpha}}{(t+1)}\right)\\ &+\frac{2ntA_{\alpha\delta}e^{-i\alpha}}{(t+1)}+\frac{2tA_{\alpha\delta}e^{-i\alpha}}{(t+1)}+\frac{2nA_{\alpha\delta}e^{-i\alpha}}{(t+1)}\Bigg)\Bigg|\left|\int_X (x)^{n-1}\, d\mu(x)\right|,\end{aligned}$$

and

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

then again since $e^{-i2\alpha}-2\delta e^{-i\alpha} = 2A_{\alpha\delta}e^{-i\alpha}-1$ and $|e^{-i\alpha}| = 1$,

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

thus

$$\begin{aligned}|a_n| &= \frac{1}{(t+1)n}\left|\left(2A_{\alpha\delta}(n-1)+t+1\right)\right|\int_X |(x)^{n-1}|\, d\mu(x)\\ &\le \frac{1}{(t+1)n}\left(2A_{\alpha\delta}(n-1)+t+1\right)\int_X |(x)^{n-1}|\, d\mu(x),\end{aligned}$$

and

$$|a_n| = \frac{1}{n}\left(\frac{2A_{\alpha\delta}(n-1)+t+1}{(t+1)}\right), \quad n = 3,5,7,\ldots$$

as required.

4. Conclusion

In conclusion, there are three purposes of this paper, which are to to define the new generalized class of tilted univalent analytic functions, to find representation theorem for this class of functions and to determine the coefficient bound for the $S^*(\delta,t)$. We believe that we have achieved all the objectives that we highlighted.

References

Darus, M. (2002). The Fekete-Szegö theorem for close-to-convex functions of the class KSH (α, β)[J]. *Acta Mathematics Academiae Paedagogicae Nyíregyháziensis*, 18, 13-18.

Duren, P. L. (1983). *Univalent Functions*. In Springer.

Goodman, A. W. (1983). *Univalent functions.* Mariner Pub. Co.

Kaharudin, N. (2011). On a class of a -close-to-convex functions. *Master Thesis, Universiti Teknologi MARA*.

Yahya, A., Soh, S., & Mohamad, D. (2013a). Coefficient bound of a generalized close-to-convex function. *International Journal of Pure and Applied Mathematics*, 83(2), 287-293.

FACTORS AFFECTING STUDENTS' ACADEMIC PERFORMANCE THROUGH ONLINE DISTANCE LEARNING

Isnewati Ab Malek[1], Amira Atiqah Mohd Nazri[2], Nur Iffa Aisha Jamshah[3], Nur Syarah Md Hashim[4], Haslinda Ab Malek[5*]

[1,3, 4,5]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Negeri Sembilan Branch, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia

[2]Downstream Digital and Innovation, Petroliam Nasional Berhad (PETRONAS) Level 25, Menara 3, Kuala Lumpur City Centre 50088 Kuala Lumpur

[*]haslinda8311@uitm.edu.my

Abstract

Nowadays, educational institutions are compelled to modify their instructional approaches to incorporate digital platforms which leading to the emergence of Online Distance Learning (ODL). The transition from traditional educational settings to flexible learning in response to the COVID-19 pandemic has presented many obstacles for both educators and learners. Various platforms, including university academic websites, Google Classroom, Microsoft Teams, Zoom, and others, are employed to facilitate the fulfilment of students' online learning obligations. However, the students faced many circumstances regarding their virtual classes such no proper devices to equip during study sessions, environment is not conducive, poor internet connection, worrying about graduating on time and future employment which affected their academic performance. Therefore, this study is conducted to determine the significant factors (family members, peer influence, financial management, study environment and online teaching) that affect the students' academic performance (CGPA). Based on the Multiple Linear Regression analysis, this study found that variable of family members (p-value = 0.027) affects significantly to the academic performance. This study can be beneficial and raises awareness to many people including family members, lecturers, counsellors, university administration and government. The academic accomplishment of students is more certain when all stakeholders fulfil their respective responsibilities.

Keywords: Academic performance, Online distance learning, Students, Multiple linear regression

1. Introduction

The Coronavirus Disease (COVID-19) pandemic has resulted in the closure of educational institutions worldwide. All students regardless of age had to experience online classes at home instead of having face to face classes According to Li and Lalani (2020), over 1.2 billion of children, teenagers and young adults are out of the classroom around the world. As a result, educational institutions are being pushed to shift their learning techniques to digital platforms, giving rise to the term of Online Distance Learning (ODL). Platforms such as the university academic website, Google Classroom, Microsoft Teams and zoom are used to cater the students to do their tasks during ODL. The process was not easy for many students especially to the ones who are less fortunate where they are not well equipped with devices such as smartphones, laptops or tablets and their study place is not conducive enough to be conducted at home with the chaos happening around them.

Moreover, a study conducted by Al-Kumaim et al. (2021) found that 69.5% of respondents representing university students in Malaysia felt overburdened through ODL. This problem arises due to the situation of poor internet connection or limited internet access during virtual classes, lack of ICT (Information Communication Technology) skills and lack of experience using online learning platforms. Furthermore, according to Chisadza et. al (2021), students' performance is related to good online access, whether via WiFi or mobile internet data. The same study also found that students who encountered difficulties during this rapid transition to online classes had poorer academic performance. The researchers suggested that the digital infrastructure should be improved and amend to the needs of learning.

The Malaysian Communications and Multimedia Commission (MCMC) revealed that broadband internet traffic soared by 23.5% during the first week of Movement Control Order (MCO) due to e-learning, work from home, etc. Many students from the Bottom Tier Income Earners (B40) depend on their hotspot usage to amend with their classes. Moreover, poor living conditions are said to be one of the difficulties that students experience when attending online classes. It is crucial for advanced education foundations to help these students, particularly with adapting to anxiety during this upsetting period.

In addition, study conducted by Arias, Swinton, & Anderson (2018) found that students' performance in academics are much more outstanding and statistically better in face-to-face classes compared to online classes for assessments such as test, final examination and assignments. This study has demonstrated that the challenges experienced by university students during ODL have a significant influence on their academic success. Therefore, the purpose of this study is to determine the factors influencing students' academic performance through ODL.

2. Literature Review

A study conducted by Son et. al (2020) has found that 82% of participants from universities in the United States agreed that their studies are badly affected by the outbreak. The participants' responses displayed that initially, they were concerned with the sudden transition from having physical classes to online classes. The university students were concerned about their progress in their tasks online such as conducting research and projects. The students also mentioned about their lack of motivation to study. These findings have shown that many students around the world regardless of their age, race, nationality and ranking of the universities are impacted from ODL method. Students are not disregarded from being affected by the pandemic especially when they have no other option than having to have online classes. With this swift change in the education field, students' academic performances are on the edge because they must go through numerous obstacles especially with studying online using their own technology equipment at home, (Al-Kumaim et al., 2021).

2.1 Family Members

During ODL, most of the students are at their home where their family lives instead of being on campus. Moreover, students must be prepared with having to do their classes at home. According to Gao et al. (2021), the process of online learning is significantly influenced by the support from the family. The researchers used a multiple mediation model which concluded that it can academically help students when the family creates a conducive environment to do online classes. Dedication and positive emotions shown by family members also aid in the students' success of academic performance. OECD (2020) supported that students' attitude towards ODL is greatly influenced by the support received by families. One of the forms of family support includes parental emotional support which is prominent in the development of positive attitude while facing ODL. Having a positive attitude during classes can maximize students' ability to do their best with the opportunities given. Unfortunately, some families are not capable in providing those support especially during the pandemic due to lack of time and having other commitments.

2.2 Peer Influence

According to Sun, Lin, and Chung (2020), due to the ODL method, students have little opportunities to interact with their peers physically. Although the existence of online platforms makes people feel more connected to each other no matter the distance, it cannot replace the face-to-face human interaction where people can share thoughts and ideas between them. It is much easier to guide and help each other with physical interaction rather than through online platforms. Peers are simply other students who understand and go through the same struggle with online learning. During this time around, it is best for them to play a part as a good support system to students as they relate to the same pressure during the pandemic outbreak. Wilczewski, Gorbaniuk, and Giuri (2021) highlights the struggles of international students. The experience of being a part of this pandemic season is challenging for this group of students as many of them did not get enough support from family and friends physically. Therefore, it can affect students' academic performance.

2.3 Financial Management

Furthermore, Dang and Bulus (2015) investigated and found that education is expensive to students who do not receive or apply any financial aid during their studies because it can arise to have financial strains to them and causes them to perform poorly in their academic. Some students are still facing financial problems like mismanaging their finance even though they received aids, loans and scholarships from numerous organisations, authorities and companies. Based on Norvilitis et al. (2006), it has shown that students with huge number of debts are corresponded to lacking information in managing their money issues and students with many financial obligations are very self-aware of their crisis could lead to face.

2.4 Study Environment

Besides that, a study space really plays a huge role for students to have their own privacy to study online. Through ODL, students should be extra cautious with their surroundings' cleanliness to avoid attracting diseases that could be fatal to them and others. Besides, Zhong, Yuan, and Fleck (2019) stated that if the study space lacks fresh air, the temperature is too high or too low, too many noises in a room, uneven lighting to study can influence the students' performance in academics during their online classes. Furthermore, students must consider using comfortable and safe furniture like getting a proper table to study and a chair to sit on when studying to improve focus in classes and during studying. According to Parvez, Rahman, and Tasnim (2019), students who sit on uncomfortable furniture could cause bad posture of their back and suffer problems like having bad backache.

2.5 Online Teaching

Learning and teaching online is not an easy task done by students, teachers and lecturers especially in this pandemic breakout. According to Lederman (2020), educators and students are put under so much pressure when they must go through online learning. Research conducted by Lim (2020) found that with many online educational platforms on the internet, some websites and applications could cause hassle and complication for the students to use them since the features were not friendly enough to use. For instance, when there are too many students in a video-conferencing classroom, there will be some troubleshooting problems that would cause some of the students with poor internet connection to be dropped from the call and causes lack of focus in class.

3. Methodology

This section discusses the sample size, data collection procedures, and data analysis methods employed in this study.

3.1 Research Design

The data was obtained through online questionnaires which determined that it is a primary data. The factors chosen specifically targeting on family members, peer influence, financial management, study environment and online teaching. The data used for this study was only obtained once from the respondents. Therefore, the single cross-sectional design is the most suitable method to be applied in collecting the data from the students.

3.2 Sampling Method

The sample method employed in this study was non-probability sampling, namely the convenience sampling method. This sampling method is very practical to apply in this study because of the less time taken to collect data and easily distributed to any students of UiTM Seremban 3.

3.3 Research Instrument

Questionnaire is designed and divided into 6 sections which are section A, B, C, D, E and F. Section A is demographic where respondents are asked about their background such as their email, gender, level of study, student ID, telephone number, age, faculty, semester, Cumulative Grade Point Average (CGPA), preference on mode of learning, and family income (RM). In section B, C, D, E and F the respondents are required to answer questions on factors that affect their academic performance during ODL among university students. These sections used Likert Scale with ten options of answers provided from “Strongly Disagree” to “Strongly Agree”.

These questionnaires were generated by adopting from much previous research regarding the factors that affect the academic performance during ODL among students around the world. The instrument used in the study was culled from the different authors. The data has been collected, organized then proceed with analyzing the data through Statistical Package for Social Sciences (SPSS).

Constructed statements that measured the research participants’ responses were adapted from several instruments and validated to test the reliability. The Cronbach’sAlpha value for each independent variable exceeded 0.6. The variable with the highest value of Cronbach’sAlpha (0.952) is study environment while online teaching has the lowest value of Cronbach’s Alpha

(0.834). Lastly, the overall reliability value of Cronbach's Alpha is 0.641, indicating that all scales of independent variables are reliable and consistent.

3.4 Population and Sample

The population for this study were all the students from Universiti Teknologi Mara (UiTM) Negeri Sembilan campus of Seremban 3. There are 5642 students in total consisted of diploma and degree students from three faculties that are available in UiTM Seremban 3 which are Faculty of Administrative Science and Policy Studies (FSPPP), Faculty of Computer and Mathematical Sciences (FSKM) and Faculty of Sports Science and Recreation (FSR). For this study, a total of 360 samples are chosen from diploma and degree students from FSPPP, FSKM and FSR in semester 2 until semester 7. The number of samples was calculated using Raosoft software. Table 1 shows the demographic profile of the respondents.

Table 1: Demographic profile of the respondents (n=360)

Characteristics	Frequency	Percentage
Gender		
Male	78	21.57
Female	282	78.43
Level of Study		
Diploma	35	9.62
Bachelor's Degree	325	90.36
Faculties		
Faculty of Sport and Recreation	22	6.12
Faculty of Administrative Science and Policy Studies	153	42.57
Faculty of Computer and Mathematical Sciences	185	51.31
Semester		
2	4	1.17
3	100	27.99
4	59	16.33
5	82	22.74
6	100	27.70
≥7	15	4.08

3.5 Theoretical Framework

Figure 1 shows that the academic performance (CGPA) which is the dependent variable in this study depends on the independent variables which are the factors that affects the academic performance through ODL among UiTM Seremban 3 students namely as family members, peer influence, financial management, study environment and online teaching.

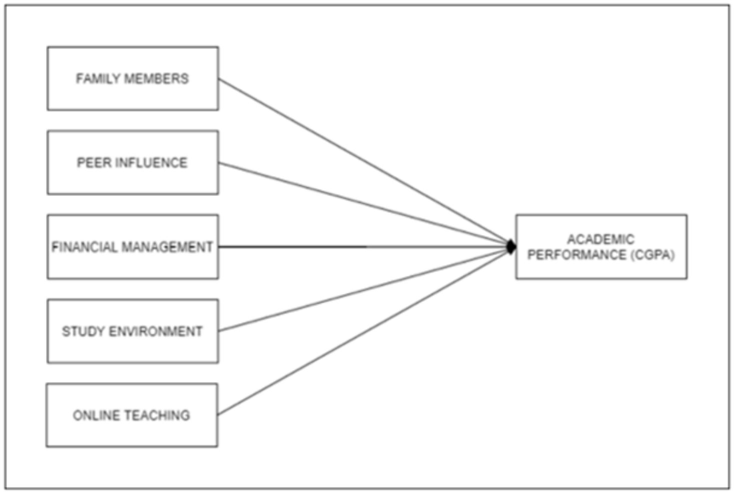


Figure 1: The theoretical framework of the study

3.6 Data Analysis

In multiple linear regression, the determination of which independent variables contribute significantly to explaining the variability in the dependent variable is a common goal. The procedure constructs a series of regression models in which variables are added or removed at each step. In general, the multiple regression equation of y on X_1, X_2 , X_3 , X_4 and X_5 is given by (1):

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \tag{1}$$

where:

y is the academic performance

β_0, β_1, β_2, β_3, β_4 and β_5 are the regression parameter or regression coefficient

X_1	:	Family Members
X_2	:	Peer Influence
X_3	:	Financial Management
X_4	:	Study Environment
X_5	:	Online Teaching
ε	:	random error

A p-value of less than 0.05 in the regression analysis indicates a significant influence of the factors identified in this study on the Students’ Academic Performance through ODL. Prior to multiple regression analysis, statistical assumptions such as linearity, homoscedasticity, independence of errors, normality of error, and independence of independent variables (no multicollinearity) were met.

4. Result and Analysis

This research was conducted to explore the most significant factor(s) that affects the academic performance during ODL among UiTM Seremban 3 students. Multiple linear regression is used to determine the significant factors consisting of family members, peer influence, financial management, study environment and online teaching that affects academic performance (CGPA).

Table 3 shows that p-value (0.034) is lower than the significant value (α = 0.05). Thus, indicating the model is significant. Table 4 are consisted of the values of R, R^2 and Durbin-Watson for the model. Since the value of R^2 is 0.035, it indicates that 3.5% of the total variation in CGPA is explained by the factors affecting students' academic performance through ODL among UiTM Seremban students, while 96.5% is explained on other factors.

Table 3: ANOVA Results

Source of Variation	Sum of Square	df	Mean Square	F	Sig.
Regression	1.592	5	0.318	2.437	0.034
Residual	44.047	337	0.131		
Total	45.640	342			

Table 4: Model Summary

R	R^2	Durbin-Watson
0.187	0.035	1.844

Table 5 shows the results of analysis by using all parameters included in the model which are family members, peer influence, financial management, study environment and online teaching as the independent variables while the dependent variable is the academic performance (CGPA) of students in UiTM Seremban 3. It is shown that there is only one significant variable which is family members (p-value = 0.027). In addition, the value of unstandardized coefficient for family members is (0.003) indicating that every one unit in total score of family members increases, the academic performance (CGPA) of UiTM Seremban 3 students increases by 0.003. This finding can be supported by research conducted by Tus (2021) where parental participation is a significant predictor of students' academic achievement, especially during this crucial time. Involvement and support from parents play a huge role in improving the students' academic performance. Therefore, it benefits the students when parents acknowledge and understand their study circumference in this pandemic. Furthermore, according to Gao et al. (2021), a happy and healthy home atmosphere can boost student academic engagement especially during this pandemic. Furthermore, a study carried out by Frawley et al. (2019), family support may influence their children's emotional experiences with learning. Therefore, this evidence supports the significance of family members affecting academic performance during this pandemic.

Table 5: Coefficients results

Model	Unstandardized coefficient		t	Sig.	Collinearity Statistics		Remarks (Hypothesis supported or not)
	B	Std error			Tolerance	VIF	
(Constant)	3.223	0.138	23.385	0.000			
Family Members	0.003	0.001	2.216	0.027	0.671	1.490	Supported
Peer Influence	0.001	0.002	0.485	0.628	0.661	1.513	Not Supported
Financial Management	-0.006	0.004	-1.537	0.125	0.954	1.048	Not Supported
Study Environment	0.000	0.001	0.123	0.902	0.597	1.676	Not Supported
Online Teaching	-0.003	0.003	-0.827	0.409	0.720	1.388	Not Supported

5. Conclusion

Multiple Linear Regression is used in determining the most significant factor affecting academic performance through ODL. Online questionnaires were distributed to 360 samples in three different faculties. The factors that contribute to the success of the academic performance of students are explained by 3.5% of the factors identified in this study. The model with all variables has fulfilled all the assumptions after running the test. The results obtained showed that there is only one variable that is significant which is family members (p-value=0.027). This study proves that the support and guidance of the parents to the students play a vital role in the success of the students. It is hoped that this study can be beneficial and raise awareness to many people including family members, lecturers, counsellors, university administration and government. The academic accomplishment of students is more certain when all stakeholders fulfil their respective responsibilities.

The future researcher can add more factors such as internet connection, intimate relationships, self-motivation, health issues and study course of study. By adding more factors, the respondents have more options of which are affecting their academic performance through ODL. Moreover, to reach a reliable result, a method of data analysis that is simultaneous and integrated is necessary. Structural Equation Modelling (SEM) is a multivariate analysis which can be applied in multi-variable and multi-relations data at the same time and is ready to test complicated relations between factors.

Acknowledgment

The researcher would like to express their sincere gratitude to all those who contributed significantly to its successful execution. First and foremost, heartfelt appreciation goes to the respondent who dedicated their time to complete the survey. Furthermore, special thanks are extended to the management of Universiti Teknologi MARA Seremban Campus for their invaluable assistance and support in facilitating the collaboration essential for this research project.

References

Al-Kumaim, N. H., Alhazmi, A. K., Mohammed, F., Gazem, N. A., Shabbir, M. S., & Fazea, Y. (2021). Exploring the Impact of the COVID-19 Pandemic on University Students'

Learning Life: An Integrated Conceptual Motivational Model for Sustainable and Healthy Online Learning. Sustainability, 13(5), 2546. https://doi.org/10.3390/su13052546

Arias, S. &, & Anderson. (2018). Online Vs. Face-to-Face: A Comparison of Student Outcomes with Random Assignment. Journal of Business Education & Scholarship of Teaching, 12(2), 1–23. https://files.eric.ed.gov/fulltext/EJ1193426.pdf

Chisadza, C., Clance, M., Mthembu, T., Nicholls, N., & Yitbarek, E. (2021). Online and face-to-face learning: Evidence from students' performance during COVID-19 pandemic. Afr Dev Rev,1-12. https://doi.org/10.1111/afdr.12520

Dang, E., & Bulus, E. (2015). The Impact of Finance on the Academic Performance of Secondary School Students in Akwanga Local Government of Nassarawa State. Nigeria. J. Of Social Sciences and Public Policy, 7(2), 44–54. https://www.cenresinjournals.com/wp-content/uploads/2020/02/Page-44-54-0520.pdf

Frawley, T., Carroll, L., Casey, M., Davies, C., Durning, J., Halligan, P., Joye, R., Redmond, C., & Fealy, G. (2019). Evaluation of a national training programme to support engagement in mental health services: Learning enablers and learning gains. Journal of Psychiatric and Mental Health Nursing, 26(9-10), 323–336. https://doi.org/10.1111/jpm.12535

Gao, H., Ou, Y., Zhang, Z., Ni, M., Zhou, X., & Liao, L. (2021). The Relationship Between Family Support and e-Learning Engagement in College Students: The Mediating Role of e-Learning Normative Consciousness and Behaviors and Self-Efficacy. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.573779.

Lederman, D. (2020) Will Shift to Remote Teaching Be Boon or Bane for Online Learning. Inside Higher Ed, 1-27.-References - Scientific Research Publishing. (n.d.). Www.scirp.org. https://www.scirp.org/(S(351jmbntvnsjt1aadkozje))/reference/referencespapers.aspx?referenceid=3023395

Li, C. and Lalani, F. (2020) The COVID-19 Pandemic Has Changed Education Forever. This Is How. https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning

Lim, M. (2020, March 20). Educating despite the Covid-19 outbreak: lessons from Singapore. Times Higher Education (THE). https://www.timeshighereducation.com/blog/educating-despite-covid-19-outbreak-lessons-singapore

Norvilitis, J. M., Merwin, M. M., Osberg, T. M., Roehling, P. V., Young, P., & Kamas, M. M. (2006). Personality Factors, Money Attitudes, Financial Knowledge, and Credit-Card Debt in College Students1. Journal of Applied Social Psychology, 36(6), 1395–1413. https://doi.org/10.1111/j.0021-9029.2006.00065.x

OECD. (2020, Sep). Strengthening online learning when schools are closed: The role of families and teachers in supporting students during the covid-19 crisis. Retrieved from https://www.oecd.org/coronavirus/policy-responses/ strengthening-online-learning-when-schools-are-closed-the-role-of-families-and-teachers-in-supporting-students-during-the-covid-19-crisis-c4ecba6c/

Parvez, M. S., Rahman, A., & Tasnim, N. (2019). Ergonomic mismatch between students anthropometry and university classroom furniture. Theoretical Issues in Ergonomics Science, 20(5), 603–631. https://doi.org/10.1080/1463922x.2019.1617909

Son, C., Hegde, S., Smith, A., Wang, X., & Sasangohar, F. (2020). Effects of COVID-19 on college students' mental health in the united states: Interview survey study. Journal of Medical Internet Research, 22(9), 1–14. https://doi.org/10.2196/21279

Sun, Y., Lin, S.-Y., & Chung, K. K. H. (2020). University Students' Perceived Peer Support and Experienced Depressive Symptoms during the COVID-19 Pandemic: The Mediating Role of Emotional Well-Being. International Journal of Environmental Research and Public Health, 17(24), 9308. https://doi.org/10.3390/ijerph17249308

Tus, J. (2021, June). (PDF) Amidst the Online Learning in the Philippines: The Parental Involvement and Its Relationship to the Student's Academic Performance. ResearchGate. https://www.researchgate.net/publication/352368179_Amidst_the_Online_Learning_in_the_Philippines_The_Parental_Invol

Wilczewski, M., Gorbaniuk, O., & Giuri, P. (2021). The Psychological and Academic Effects of Studying From the Home and Host Country During the COVID-19 Pandemic. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.644096

Zhong, L., Yuan, J., & Fleck, B. (2019). Indoor Environmental Quality Evaluation of Lecture Classrooms in an Institutional Building in a Cold Climate. Sustainability, 11(23), 6591. https://doi.org/10.3390/su11236591

FINANCIAL BEHAVIOUR OF YOUNG ADULTS: THE INFLUENCE OF SOCIOECONOMIC STATUS, FINANCIAL LITERACY AND PARENTAL FINANCIAL SOCIALIZATION

Muhammad Amirual Azlam[1], Alia Nadhirah Abd Malek[2], Faqihah Syazwani Abdul Mutaliff [1] and Nurdia Azlin Ghazali [3,*]

[1,2,3]College of Computing, Informatics and Mathematics
Universiti Teknologi MARA Cawangan Negeri Sembilan
Seremban Campus, 70300 Negeri Sembilan
*nurdia040@uitm.edu.my

Abstract

This study aimed to analyze the factor influencing financial behavior among young working adults at Universiti Teknologi Mara (UiTM) Shah Alam, Selangor. The purpose was to determine whether these young adults exhibit good financial behavior and to identify which factors; socioeconomic status (academic qualification, income, number of dependents), financial literacy and parental financial socializations significantly impact their financial behavior. The sample comprised 210 respondents including employees of the institution. Data was collected through a self-administrated online questionnaire. The dependent variable, financial behavior was categorized into three levels: low, moderate and high. Descriptive statistics and Ordinal Logistic Regression were employed to assess the factors affecting financial behavior. The finding indicates that most young working adults do practice good financial behavior. Additionally, the results reveal that academic qualification, income and number of dependents do not significantly influence financial behavior. In contrast, financial literacy and parental financial socialization were identified as significant factors of financial behavior among young working adults.

Keywords: Financial behavior, young adults, financial literacy, parental financial socialization, socioeconomic status

1. Introduction

Financial stability plays an important role in everyone's life as it represents the state of financial health that results from a good financial behaviour. When a person is financially stable, they can consistently meet their financial obligation consistently without experiencing stress. Financial behavior is one of the key contributors to having better financial stability and generally people with a high degree of financial behavior tend to be in a better financial position (Rahman et al., 2021). Financial behavior can be defined as any human behavior related to managing money (Xiao, 2008). Financial behavior also can be seen from how an individual manages its financial resources, including cash, debt, savings and other expenses (Hasibuan et al, 2018). However, millennials are increasing experiencing financial stress with many of them living beyond their means, being trapped in emotional spending and being on the edge of financial instability (Azmi and Madden, 2015).

Everyone manages their money differently and effective money management requires effort, time and consistency. Controlling spending habits, especially for young adults who may struggle with shopping impulses can be challenging. Young adults may either handle their financial behavior successfully or face serious issues such as bankruptcy, particularly as they manage money independently for the first time. According to Mohamad (2020), financial literacy is alarmingly low among young adults in Malaysia. The article notes that between 2015 and 2019, approximately 84,805 Malaysians were declared bankrupt. The insolvency department stated that, of these, 26% were young adults under the age of 34. Additionally, 47% of Malaysian youth have high credit card debt, largely due to difficulties in managing instalment purchases, personal loans and credit card debt.

Financial literacy refers to the ability to understand and apply various financial concepts and techniques, such as personal financial management, budgeting, and investing. It is crucial for an individual to manage their finances because it provides skill and knowledge to make financial decisions to maintain financial stability to achieve good financial behaviors. According to Herawati et al (2018) financial literacy and socioeconomic status has a significant impact on respondents' financial behavior. In the study by de Bassa Scheresberg (2013) found that individuals with lower incomes or less education tend to have particularly low financial literacy. However, even a high level of education does not necessarily guarantee financial literacy. While individuals with college degrees generally demonstrate better financial management, financial literacy remains relatively poor even among those with a high degree of education.

According to Danes (1994) as referenced in Jung (2021), financial socialization is the process through which individuals acquire values, knowledge, norms, attitudes, and behaviors that promote financial well-being and viability. Financial socialization agents include family, parents, education, peers, and media. Among these, parental financial socialization is particularly influential in shaping children's financial behavior. Shim et al. (2010) suggest that parents and family play a crucial role in children's acquisition of financial knowledge and the development of their future financial behaviors. Parents are often the primary role models, and the first social unit children encounter Jung (2021), influencing their financial attitudes through daily routines related to managing, spending, saving, and investing. That behaviors can be formed by observing family members' daily routines in relation to financial tasks such as managing, spending, saving and investing, as well as through purposeful forms of socialization in the family such as giving advice, opening accounts, and participating in the management of household income (Chowa and Despard, 2014).

Socioeconomic status refers to the social standing or class of an individual or group which can be measured according to the combination of family background, education, income or number of dependents. According to Herawati et al. (2018) a family's social condition plays a role in the students' financial behavior. Different levels of socioeconomic status will affect the emergence of differences in perception towards physical objects or behavioral objects. This resulting in various attitude based on levels of socioeconomic. De Bassa Scheresberg (2013) also found that income and education are linked to financial outcomes.

The potential for bankruptcy poses significant concerns for young adults' futures. Young adults are at critical stage where the begin to make significant financial decision independently, without proper financial planning, young adults who face unforeseen expenses are at increased risk of financial stability without financial planning and practicing a good financial behavior. By targeting young adults in financial behavior studies, this study can identify factors influencing financial behavior. Thus, addressing financial behavior is crucial to prevent early financial difficulties and ensuring better financial management practices among young adults. Hence, the following hypotheses were developed:

H_1: There is a significant influence between academic qualification and financial behaviour.

H_2: There is a significant influence between income and financial behaviour.

H_3: There is a significant influence between number of dependent and financial behaviour.

H_4: There is a significant influence between financial literacy and financial behaviour.

H_5: There is a significant influence between parental financial socialization and financial behaviour.

2. Methodology

2.1. Research Design

This cross-sectional study aims to provide a comprehensive understanding of the factors influencing financial behavior among young adults. This study used quantitative analysis to assess the relationships between academic qualification, income, number of dependents, financial literacy and parental financial socialization as predictors of financial behavior. This study involved a sample of 210 young adults currently employed at a tertiary institution in Selangor. The target population consisted of young working adults in the age range of 18 to 40 years. Figure 1 illustrates the research framework of this study where the independent variables are academic qualification, income, number of dependents, financial literacy and parental financial socialization are analyzed as factors contributing to the financial behavior.

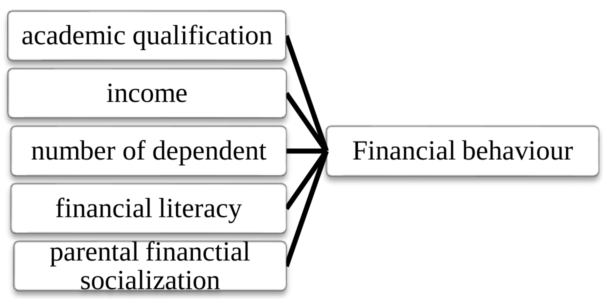


Figure 1: Research framework of this study

2.2. Research Instrument

Data was collected using a structured questionnaire administered through a survey. The questionnaire is divided into three sections: Section A contains questions on demographic information of respondents which include respondents' age, gender and marital status. This section also includes questions on demographic factors related to financial behavior which are academic qualification, income and number of dependents. These factors were assessed using multiple-choice questions. Section B consists of factors of financial behavior which are financial literacy and parental financial socialization. For factor financial literacy, questions were expressed as multiple-choice items. Responses were recorded with a score of 1 for each correct answer and a score of 0 for incorrect answers. The accumulated scores were calculated to determine the extent to which financial literacy affects financial behavior among young adults. For parental financial socialization, the questions were measured using a 5-points Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Section C consists of items related to financial behavior. This section focuses on cash management, credit card management and saving behavior. Respondents rated their financial behavior from 1 (strongly disagree) to 5 (strongly agree). Each response was scored based on the option selected for each question. The questionnaires items related to financial literacy were adapted from Stella et al. (2020) while those on parental financial socialization were from LeBaron-Black et al. (2021) and financial behavior were adapted from Hogarth et al., (2003). Table 1 summarizes the scores and level of financial behavior.

Table 1: Levels of Financial Behavior Based on Scores

Levels of financial behavior	Scores
Low	Below than 30%
Moderate	30% - 75%
High	More than 75%

2.3. Measurement Reliability

To ensure the collected data is reliable and can produce precise and constant result, a reliability test was conducted. The degree of reliability was determined by using Cronbach's alpha where a value more than 0.7 indicates a higher level of internal consistency reliability Sekaran and Bougie (2010).

2.4. Data Analysis

Descriptive statistics were employed to summaries the information about the respondent and to explore whether the young adult practices good financial behavior in terms of saving, cash and credit card management. Next, ordinal logistic regression was employed to explore the relationship between independent variables to determine the influence of these factors on students' dependent variable that has three categories (low, moderate or high financial behavior). The Wald test was used to assess the significance of individual coefficients or factors within the model. The significance of the Wald statistic can be seen from the p-value for each variable which must be less than alpha equal to 0.05 to indicate the importance of the predictor variables in the model. High values of the Wald statistic also will show that the corresponding predictor variable is significant. If the null hypothesis is rejected, the independent variables will be significant to determine that the independent variables influence the financial behavior.

3. Results and Discussions

3.1. Reliability Result

The results of reliability test, as shown in Table 2, indicate that all variables are highly reliable and dependable for further analysis as the Cronbach's Alpha values are greater than 0.7.

Table 2: Reliability Results for Variables

Variable	Cronbach's Alpha
Financial behaviour	0.896
Parental Financial Socialization	0.909

3.2. Demographic Profile

The descriptive statistics for 210 respondents are shown in Table 3. The data reveals that there are 57.7% of the respondents were female and 44.3% indicating a higher proportion of female participants compared to males. About 54.3% of respondents under 25 years, 17.1% were aged 26 – 30 years, 16.7% were aged 36 – 40 years. It also can be observed that, majority of the respondents had a degree with 43.8%. Regarding income, 20.5% of the respondents earned less than RM1000, 34.8% earned income between RM1001 and RM2500, 22.4% earned income between RM2501 and RM3500, 11% earned income between RM3501 and RM5000 and 11.4% earned more than RM5000. For the number of dependents, 64.3% of respondents had 3-6 dependents, 23.3% had fewer than 3 dependents and 12.4% had more than 6 dependents.

Table 3: Summary of Demographic Information of students

Variable	Category	Frequency	Percentage
Gender	Male	93	44.3
	Female	117	55.7
Semester	Below 25	114	54.3
	26 – 30	36	17.1
	31 – 35	35	16.7
	36 – 40	25	11.9
Academic Qualification	PMR	2	1.0
	SPM	27	12.9
	Diploma	66	31.4
	Bachelor's degree	92	43.8
	Master and PhD	23	11.0
Total income	Less than RM1000	43	20.5
	RM1001 – RM2500	73	34.8
	RM2501 – RM 3500	47	22.4
	RM3501 – RM 5000	23	11.0
	RM5000 and above	24	11.4
Number of Dependents	Less than 3	49	23.3
	3 – 6	135	64.3
	More than 6	26	12.4

Table 4 shows the descriptive statistics of mean score for factors financial literacy and parental financial socialization among the 210 respondents. For the variable Financial Literacy, scores ranged from a minimum of 0 to a maximum of 6 indicating a spectrum of knowledge levels from no financial knowledge to high financial knowledge. The mean score of Financial Literacy is 3.65 suggesting, most of the respondents do have knowledge about Financial Literacy such as respondents can handle financial situations that require them to do so more effectively. According to Stella et al. (2020), a score between 3 and 4 indicates good financial literacy. For the Parental Financial Socialization, scores ranged from a minimum of 8 to a maximum of 40. The average score was 29.99, indicating a good Parental Financial Socialization among respondent as indicated by LeBaron-Black et al. (2021), scores between 25 and 30 indicate good parental financial socialization.

Table 4: Descriptive Statistics of Factors of Financial behavior

Variable	Mean Score	Variable
Financial literacy	3.65	Financial literacy
Parental Financial Socialization	29.99	Parental Financial Socialization

From Figure 2, it is observed that more than half of the total respondents from the tertiary institution fall into the high financial behavior group with 54.3%. The remaining respondents are nearly equally divided between the low and moderate financial behaviors groups, with 23.3% in the low group and 22.4% in the moderate group. This result shows that most young working adults at the tertiary institution demonstrate good financial behaviors and possess substantial knowledge on finances. However, the rest of them were divided into low and moderate financial behaviors categories. The financial behavior among young working adults is concerning with 23.3% who have poor financial behavior in their daily life.

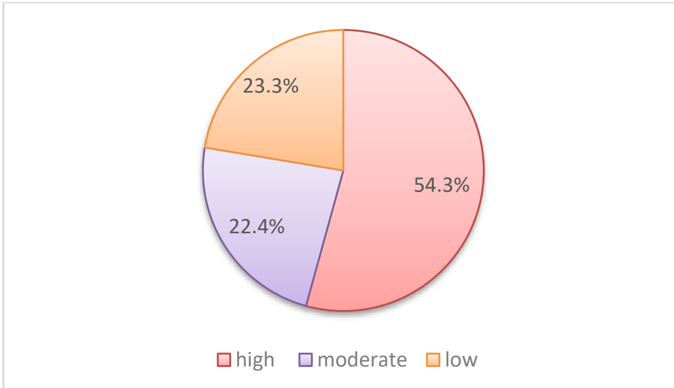


Figure 2: Financial Behavior of Respondents

3.3. Factors Influencing Financial Behavior

Table 5 presents the results of ordinal logistic regression model. Financial literacy and parental financial socialization were statistically significant at the 5% significant level suggesting that these factors significantly influence the financial behavior of young working adults. For each one unit increase in Financial Literacy, there is a predicted increase of 0.325 in the log odds of individual being at a higher level of Financial Behavior. This indicates that individuals with higher Financial Literacy are more likely to achieve better financial behavior. Thus, when individuals get more knowledge about financial concepts, they will be able to make better financial decisions. This finding aligns with Herawati et al. (2018) who state that financial literacy has a significant impact on financial behavior. Parental Financial Socialization also was a significant factor as indicated by its p-value of 0.000. For each one unit increase in Parental Financial Socialization, there is a predicted increase of 0.072 in the log odds of being in a higher level of the Financial Behavior. This suggests that higher scores in parental financial socialization are associated with better financial behavior. However, socioeconomic status such as academic qualification, income and number of dependents have no influence on the financial behavior of young working adults.

Table 5: Model Summary of Ordinal Logistic Regression Analysis model

Variables	Categories	Estimate	Significant
Financial behavior	Low	1.833	0.071
	Moderate	3.024	0.003
Academic Qualification	PMR	18.805	-
	SPM	0.069	0.921
	Diploma	0.114	0.849
	Bachelor’s degree	0.027	0.962
	Master and PhD	0	-
Income	Less than RM1000	-0.226	-
	RM1001 – RM2500	-0.705	0.921
	RM2501 – RM 3500	-0.509	0.849
	RM3501 – RM 5000	0.210	0.962
	RM5000 and above	0	-
Number of dependents	Less than 3	0.744	0.144
	3 – 6	0.086	0.843
	More than 6	0	-
Financial literacy		0.325	0.003
Parental financial socialization		0.072	0.000

4. Conclusion

The main objective of this study is to analyze the factors influencing financial behavior among young working adults in Universiti Teknologi MARA (UiTM) Shah Alam. This study employs ordinal regression models to account for the ordinal term of the dependent variable. Findings reveal that the financial behavior of young working adults is not significantly associated with academic qualification, personal monthly income and number of dependents. However, literacy and parental financial socialization are significantly associated with financial behavior among young working adults. In conclusion, focusing on financial literacy and parental financial socialization is crucial for improving financial outcomes among young adults. Increasing awareness and education about financial management across different age groups can contribute to better to secure the best financial life planning and stability. Future research should look at the gender differences in financial knowledge to identify whether there is potential dissimilarity. Besides that, expanding the study to include a larger population including people across various state in Malaysia could provide insights into regional difference in financial behavior.

Acknowledgment

The author would like to extend sincere gratitude to the research team member for their dedication, efforts, and invaluable contributions to this study.

References

Azmi, Wan & Madden, Raymond. (2015). Finance Matters: Understanding Gen Y Bridging the Knowledge Gap of Malaysia's Millennials.

Chowa, G. A., and Despard, M. R. (2014). The influence of parental financial socialization on youth’s financial behavior: Evidence from ghana. *Journal of Family and Economic Issues*, 35(3), 376–389.

Danes, S. M. (1994). Parental perceptions of children's financial socialization. *Financial Counseling and Planning*, **5**, 127–149.

de Bassa Scheresberg, C. (2013). Financial literacy and financial behavior among young adults: Evidence and implications. Numeracy, 6(2), 5.

Hasibuan, B., Lubis, Y. and Walad, H. R. (2018). Financial Literacy and Financial Behavior as a Measure of Financial Satisfaction. 10.2991/ebic-17.2018.79.

Hogarth, J. M., Beverly, S. G., and Hilgert, M. (2003). Patterns of financial behaviors: Implications for community educators and policy makers.

Herawati, T., Candiasa, I., Yadnyana, I and Suharsono. (2018). Factors That Influence Financial Behavior Among Accounting Students in Bali. *International Journal of Business Administration*. 9. 30. 10.5430/ijba.v9n3p30

Mohamad, A. (2020) Financial literacy among youth alarmingly low. New Straits Times

Jung, E. (2021). Examining the relationships among parental financial socialization, financial capability, financial behavior, and financial well-being in lower-income and higher-income young adults (Unpublished doctoral dissertation). University of Kansas.

LeBaron-Black, A. B., Curran, M. A., Hill, E. J., Freeh, M. E., Toomey, R. B., and Speirs, K. E. (2021). Parent financial socialization scale: Development and preliminary validation. Journal of Family Psychology.

Rahman, M., Isa, C.R., Masud, M.M., Sarker, M. and Chodhury, N. T. (2021). The role of financial behavior, financial literacy, and financial stress in explaining the financial well-being of B40 group in Malaysia. *Futur Bus J* **7**, 52 (2021). https://doi.org/10.1186/s43093-021-00099-0

Sekaran, U., & Bougie, R. (2010). Research methods for business: A skill-building approach (5th ed.). Haddington: John Wiley & Sons.

Shim, S., Barber, B. L., Card, N. A., Xiao, J. J., and Serido, J. (2010). Financial socialization of first-year college students: The roles of parents, work, and education. *Journal of youth and adolescence*, 39(12), 1457–1470.

Stella, G. P., Filotto, U., and Cervellati, E. M. (2020). A proposal for a new financial literacy questionnaire. *International Journal of Business and Management*, 15(2).

Xiao, J. J. (2008). Applying behavior theories to financial behavior. *In Handbook of consumer finance research* (pp. 69–81). Springer.

IMPROVING MORTALITY FORECASTING: INTEGRATING THE LEE-CARTER MODEL WITH NEURAL NETWORKS

Hana Natasya Abd Hamid[1], Khairunnisa Khairul Hizam[2] and Nurul Aityqah Yaacob[3,*]
[1,2]Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan, Malaysia, [3]Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Kuala Pilah, Negeri Sembilan, Malaysia
[*]corresponding author: aityqah@uitm.edu.my

Abstract

This research integrates the Lee-Carter (LC) model with neural network (NN) methods to enhance mortality forecasting. The LC model, widely used for demographic forecasting, has limitations in capturing complex and nonlinear mortality trends. To address these limitations, we incorporate NN methods, namely a multilayer feed-forward neural network (MFFNN), to identify patterns within mortality data. The study evaluates the performance of the LC and LC-NN models across five countries: Germany, Japan, Malaysia, South Korea, and the United States of America (USA). Findings indicate that the LC-NN model outperforms the LC model, as demonstrated by lower Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) values. This integration significantly improves forecasting accuracy, providing more reliable insights into future mortality trends. The results have significant implications for public health planning and policymaking, offering a robust tool for forecasting demographic changes and their impact on healthcare systems. This integration advances mortality forecasting, leading to better public health outcomes.

Keywords: Mortality forecasting, Lee-Carter Model, Neural network, Multilayer feed-forward network, Public health

1. Introduction

Over the past two centuries, there has been a notable and constant increase in the average life expectancy, with various countries exhibiting consistent improvements in this metric (Torri and Vaupel, 2012). The evaluation of life expectancy and mortality has been a focal point of interest in actuarial and social sciences, leading to the development of numerous models to accurately depict mortality rates (Alho, 1990). Mortality rates in a country are significantly influenced by its growth rate, as well as factors such as medical advancements and social conditions (Baran et al., 2007). These rates serve as reliable indicators of quality of life and provide indirect insights into the prevailing socioeconomic conditions. Moreover, a comprehensive analysis of past mortality trends, encompassing many causes of death, offers significant insight into the determinants contributing to the progressive escalation of mortality rates over successive periods (Zhang et al., 2023). Forecasting future mortality rates is crucial for policy formulation and planning, with the Lee-Carter (LC) model emerging as a prominent tool due to its stochastic approach and widespread applicability (Baran et al., 2007). Introduced in 1992 by Ronald Lee and Lawrence Carter, the LC model used a base model of age-specific death rates with a time series model, known as autoregressive integrated moving average (ARIMA), to effectively forecast mortality rates (Lee and Carter, 1992). The model parameters are computed using Singular Value Decomposition (SVD) and applied to the initial LC model established by Lee and Carter. After obtaining the parameters, the variable k_t will be forecasted with the help of ARIMA. The model includes two sets of age-specific constants: b_x, which reflect age-specific constants, and a_x, which characterise the general mortality pattern by age. In the original LC model, Lee and Carter (1992) proposed the standard univariate ARIMA(0,1,0) time series model which is a random walk with drift model for forecasting future mortality rates. Despite the robustness and widespread applicability of the LC model, it faces criticisms regarding its long-term stability and sensitivity to structural changes over time, which can lead to significant forecast errors if not accurately accounted for (Lee and Miller, 2001; Koissi et al., 2006). ARIMA is adept at capturing temporal dependencies and trends in mortality data, enhancing the LC model's capacity for short-term mortality forecasting (Hong et al., 2021). However, challenges such as overfitting and selecting appropriate ARIMA parameters remain, particularly in cases of sparse data or unstable mortality trends (Nigri et al., 2019).

This research, therefore, seeks to fill this gap by combining the original LC model with NNs to forecast future changes in the parameter k_t, addressing the limitations observed in ARIMA processes. By leveraging NNs to capture complex mortality trends while utilising the LC model's temporal trend modelling capability, this study aims to create a more accurate and interpretable mortality forecasting model. This integration promises to enhance both the LC model's performance and the precision of mortality forecasts by mitigating biases and incorporating structural modifications. In recent years, various researchers have proposed using machine learning (ML) algorithms to forecast death rates for populations worldwide (Hong et al., 2021). Neural Networks (NNs), in particular, have been suggested for forecasting and simulating death rates due to their ability to identify latent time frame processes and directly forecast mortality (Hainaut, 2018). ML approaches have been utilised to improve the estimation of log mortality rates within the LC framework, contributing to more robust mortality forecasts (Deprez et al., 2017; Levantesi and Pizzorusso, 2019; Richman and Wüthrich, 2018). Despite scepticism in the demographic field, recent research demonstrates the potential of ML and NNs in mortality forecasting. Hainaut (2018) employed NNs to identify hidden time processes and forecast mortality directly, while Deprez et al. (2017) enhanced mortality rate estimation accuracy using ML techniques. Richman and Wuthrich (2019) automated the structure determination of the LC model using NNs, outperforming conventional models in out-of-sample forecasts. Additionally, Hong et al. (2021) proposed a hybrid model combining the LC model, random forest (RF), and artificial neural networks (ANN), demonstrating superior mortality rate forecasting performance.

The paper is organized as follows: Section 2 presents the data and methods used in the paper. Section 3 discusses the results. Finally, Section 4 provides a summary of the conclusions drawn and outlines potential future research directions.

2. Research Method

The mortality data for five countries, which are Germany, Japan, Malaysia, South Korea, and the United States of America (USA) were sourced from the Human Mortality Database (HMD), while data for Malaysia were obtained from the Department of Statistics Malaysia (DOSM) for use in this study. The dataset comprises mortality rates for the general populace, segmented by single years of age up to 95 years for Germany, Japan, South Korea, and the USA, with Malaysia extending up to 85 years and grouped by age. Lee and Carter (1992) computed the average of $\ln(m_{x,t})$ over time to estimate a_x, then used Singular Value Decomposition (SVD) on the centred data to estimate parameters b_x, and k_t. The study commenced with a time series extracted from k_t. The data sets were split into training and testing periods using an 80:20 train-test ratio. Table 1 details the total years covered in each country's data along with the respective years allocated for training and testing sets during the analysis.

Table 1: Total, training and testing set years by country

Country	Total years	Training set years	Testing set years
Germany	1990 – 2020	1990 – 2014	2015 – 2020
Japan	1947 – 2022	1947 – 2007	2008 – 2022
Malaysia	1991 – 2015	1991 – 2010	2011 – 2015
South Korea	2003 – 2020	2003 – 2017	2018 – 2020
USA	1950 – 2021	1950 – 2006	2007 – 2021

2.1. *Lee-Carter Model*

The LC model, introduced by Lee and Carter (1992), is a statistical time series model designed to extrapolate mortality trends and age patterns. The model is formulated using the overall mortality rate (denoted as $(m_{x,t})$). The LC model is articulated as follows:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}; x = 1, \ldots, \omega \tag{2.1}$$

where a_x:

$$a_x = \frac{\sum_{t=1}^{n} \ln m_{x,t}}{n} \tag{2.2}$$

is a set of age-specific constants describing the general pattern of mortality by age (Hong et al., 2021). b_x is a set of age-specific constants describing the rates that decline rapidly and which rates decline slowly in response to changes in k, $\varepsilon(x,t)$ is the residual at age x in year t; and $m_{x,t}$ is the central death rate at age x in year t (Booth et al., 2002). The age-specific historical influences are reflected in $\varepsilon(x,t)$, which is not entirely represented by the $N(0,\sigma^2)$ distribution-following, independently and identically distributed model. The parameters in equation (2.1) are estimated using a two-stage method by imposing the following restrictions:

$$\sum_t k_t = 0 \quad \text{and} \quad \sum_x b_x = 1, \tag{2.3}$$

to identify a single solution to the model's set of equations, use the matrix of centered age profiles, $\ln(m_{x,t}) - a_x$, which subjected to the Singular Value Decomposition (SVD) technique for initial calculation of the parameters b_x and k_t (Yaacob et al., 2021).

$$b_x = \frac{1}{\sum_x u_{x,1}} (u_{1,1}, u_{2,1}, \ldots, u_{x,1})^n \tag{2.4}$$

$$k_t = \sigma_1 \times (v_{1,1}, v_{2,1}, \ldots, v_{t,1}) \tag{2.5}$$

The basic mortality rates, $\hat{m}_{x,t} = \frac{D_{x,t}}{E_{x,t}}$ are fitted to the model in equation (2.1), where $D_{x,t} > 0$ indicates the number of deaths of age x at time t, and $E_{x,t}$ is the corresponding central exposure of age at time t. In order to compare actual and forecast mortality, a second stage estimate of k_t is found once b_x and k_t are estimated by satisfying equation (2.2). This ensures that the total expected deaths for each t and the total actual deaths are equal. Thus, the estimates of the parameters meet:

$$\sum_{x=x_1}^{\omega} D_{x,t} = \sum_{x=x_1}^{\omega} E_{x,t} \left[\exp\left(\hat{a}_x + \hat{b}_x \hat{k}_t\right)\right] \forall t \tag{2.6}$$

By counterbalancing the effect of adopting a log transformation of the mortality rates, this adjustment provides more weight to high rates. Using future extrapolation values, i.e., k_{t+n}, an appropriate time series was fitted to k_t to forecast mortality rates. Therefore, the forecasted mortality rate would be:

$$\ln(m_{x,t}) = a_x + b_x k_{t+n} \tag{2.7}$$

where a_x and b_x are fixed (Yaacob et al., 2021). The ARIMA time series models are used to extrapolate the adjusted k_t. Lee and Carter (1992) used a random walk with drift model, which can be expressed as:

$$k_t = k_{t-1} + d + e_t \tag{2.8}$$

where d represents the annual change in k that is constant, and e denotes uncorrelated errors. The uncertainty correlated with a one-year forecast is approximated by the sum of the standard errors in d and e_t. This is used to generate forecasted intervals based on probabilities for the forecasted values of k_t. Age-specific mortality rates are forecasted through the utilisation of extrapolated k_t and fixed a_x and b_x. The fitted rates in this instance are the jump-off rates (i.e., the rates in the final year of the fitting period or jump-off year).

2.2. *Neural Network*

The NN becomes non-linear once we include an intermediate layer that contains hidden neurons. A simple example is shown in Figure 1. The parameters b_1, b_2, b_3 and $w_{1,1}, \ldots, w_{4,3}$ from the data are "learned" (or estimated); the objective is to minimize a function in a high-dimensional space that calculates the difference between the anticipated values ϕy and the actual values y̌. The value of y̌ is influenced by the matrices of the weights. Weight values are frequently limited to prevent them from becoming extremely huge. The "decay parameter" is a limit on the weights; it is frequently designed to have a value of 0.1. Starting with random values, the weights are afterward modified based on the observed data. Considering this, the forecasts generated by an NN contain an amount of randomness. Therefore, the network is typically trained on multiple occasions from separate random initial states, with the average outcomes being computed.

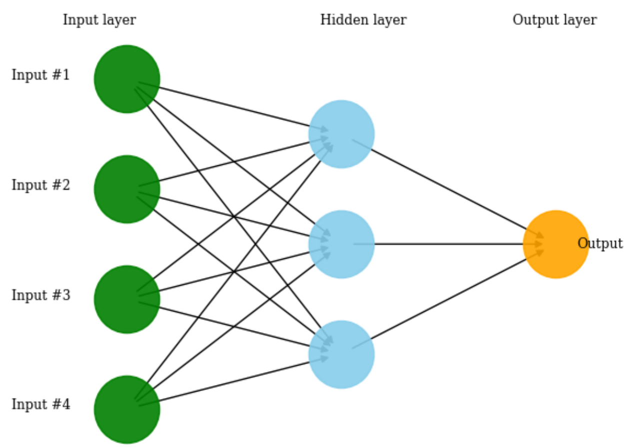


Figure 1: A NN with four inputs and one hidden layer with three hidden neurons (Hyndman and Athanasopoulos, 2021)

In Figure 1 the diagram refers to a multilayer feed-forward network in which each layer of nodes grows inputs from the past layers. The outputs of the nodes in a given layer serve as the inputs for the following layer. The inputs to each node are combined by a weighted linear combination. Afterward, the outcome is altered by a non-linear function before it is delivered as an output. This paper will employ a feedforward neural network with a multi-layer net architecture, including a single layer in the hidden layer. Following this, the backpropagation algorithm is used for training, and the sigmoid function is used for activation. A synopsis of the neural network architecture that will be implemented in this paper. The backpropagation algorithm comprised the subsequent steps, as described by Safitri et al. (2018):

i. The determination of the input nodes' quantity will be accomplished through one of the following methods: evaluation of the correlation value between variables based on relationship proximity; device-generated description of the correlogram R (Hong et al., 2021).

ii. The dataset will be split into two distinct sets: training data and testing data. The training data will be used for the purpose of learning, whereas the testing data will be employed to validate the model that was acquired, specifically to determine whether it adequately describes the existing data. Following that, normalisation of the data will occur. Following that, early initialization will be performed on all weights and bias values using a random number range of 0 to 1, and the activation function will be determined.

iii. Subsequently, input values are entered, and each input is multiplied by its corresponding weight and added together; this process advances the input to the concealed layer. The formula for this operation is as follows:

$$z_j = b_j + \sum_{i=1}^{4} w_{i,j} x_i \tag{2.9}$$

After obtaining the z_j value for each node in the hidden layer, compute the hidden layer's output signal as follows: The expression $z_j = f(z_{\text{net}j})$ denotes f, the predefined activation function.

iv. Execute step 3 in proportion to the quantity of hidden layers employed.

v. Following this, the value of the concealed layer's output will be obtained. The output of the concealed layer is transmitted to the output layer via the formula below:

$$Y_{\text{net}_k} = w_{k0} + \sum_{j=1}^{p} z_j w_{kj} \tag{2.10}$$

vi. Both the actual output and the expected output will subsequently be compared. If deemed unsuitable or if the error is substantial, weight adjustments will be applied to each layer. Weight correction may be performed using either approach. It may begin with the weight between the hidden output layers, or it may be executed from behind. The formula for modifying the weights and biases between the hidden and output layers is as follows:

$$w_{kj}^{(\text{new})} = w_{kj}^{(\text{existing})} + \Delta w_{kj} \tag{2.11}$$

$\Delta w_{kj} = \alpha \delta_k z_j$, $j \neq 0$ is correction formula for weight value and $\Delta w_{k0} = \alpha \delta_k$ is correction formula for bias value, where δ_k was obtained from formula presented below:

$$\delta_k = (t_k - Y_k) \cdot f'(Y_{\text{net}_k}) \tag{2.12}$$

vii. Then value of weight and bias between the input layer and hidden layer will be renewed with formula as follows:

$$v_{ji}^{(\text{new})} = v_{ji}^{(\text{existing})} + \Delta v_{ji} \tag{2.13}$$

$\Delta v_{ji} = \alpha \delta_j x_i$, $i \neq 0$ is correction formula for weight value, and $\Delta v_{j0} = \alpha \delta_j$ is correction formula for bias value, where δ_j was obtained from formula presented below:

$$\delta_j = (\delta_{\text{net}_j} \cdot f')(Z_{\text{net}_j}) \tag{2.14}$$

$$\delta_{\text{net}_j} = \sum_{k=1}^{m} \delta_k w_{kj} \tag{2.15}$$

This procedure is iterated until the error value is minimised to its minimum. Following this, the model with the most accurate estimation weight will be obtained for subsequent forecasting attempts (Safitri et al., 2018).

Once the training procedure is concluded, it is possible to proceed with the forecasting of the k_t values. To renormalize the forecasted k_t values, which have been normalised, the product of the differences between the minimum and maximum k_t values of the data must be applied to the k_t values. The mortality rates are then calculated using the denormalized k_t values and equation (2.1) (Hong et al., 2021).

2.3. *Evaluation Metrics*

The evaluation of forecast accuracy involves comparing forecasted values with actual values, typically done by calculating the forecast error as the difference between the forecasted and actual values. To gauge the accuracy of forecasted mortality rates, commonly used metrics such as MAPE and RMSE are employed (Ibrahim et al., 2021). In this paper, we have computed the MAPE and RMSE for each model $m_{x,t}$ to determine and compare their accuracy in forecasting mortality rates. As MAPE and RMSE decrease, the accuracy of the constructed model increases. The degree of similarity between a measured or predicted value and the genuine or accepted value of the quantity being measured or predicted is the accuracy measure. Its significance in forecasting is that it has an immediate influence on the dependability, credibility, and efficacy of assessments or prognostications. By assessing the degree to which a measurement corresponds to the actual quantity it intends to represent, accuracy measures guarantee the reliability of experimental, observational, or predicted data for the purposes of analysis, decision-making, and comparison with other datasets.

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i - P_i|}{Y_i} \times 100 \tag{2.16}$$

$$\text{RMSE} = \left[\frac{1}{N}\sum_{i=1}^{N}(P_i - Y_i)^2\right]^{\frac{1}{2}} \tag{2.17}$$

where N is the number of data, Y_i is an actual value, P_i is the forecasted value of the i^{th} data obtained.

3. Result and Discussion

This section presents the results and discussions derived from the death rate forecasting analysis using two distinct models: the LC model and the integrated LC-NN model. This chapter aims to evaluate the performance and effectiveness of these models in forecasting death rates across the five countries studied. The analysis is based on a thorough examination of observed death rates, forecasted values, and performance measures such as mean absolute percentage error (MAPE) and root mean square error (RMSE). At the outset, the chapter reviews the trends in the log death rates during different periods and among several age groups. These are critical in the basic comprehension of the essentials of the dataset before examining the difference between the forecasted and the actual death rates. After the forecasted and actual log death rates are presented, the chapter then measures the performance of the LC model and the integrated LC-NN model.

3.1. *Comparison of Forecasted and Actual Death Rates*

Figures 2–6 were derived from obserned using R-Studio, showcase the comparison between actual and forecasted death rates for total populations in five countries. Circles denote observed rates, while red (NN - Neural Network) and blue (LC – Lee-Carter) lines represent forecasted trends. These visualisations provide crucial insights into the predictive performance of these models across varied demographic scenarios.

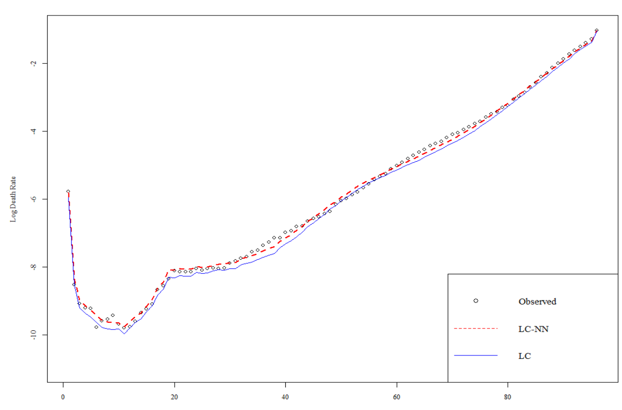


Figure 2: Observed and Forecasted Log of Death Rates for Germany in 2020

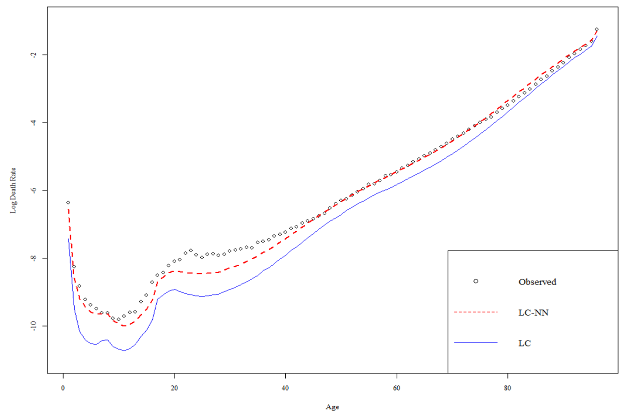


Figure 3: Observed and Forecasted Log of Death Rates for Japan in 2022

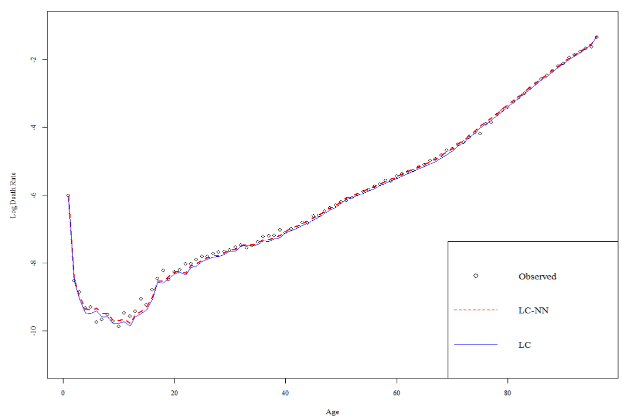


Figure 4: Observed and Forecasted Log of Death Rates for South Korea in 2020

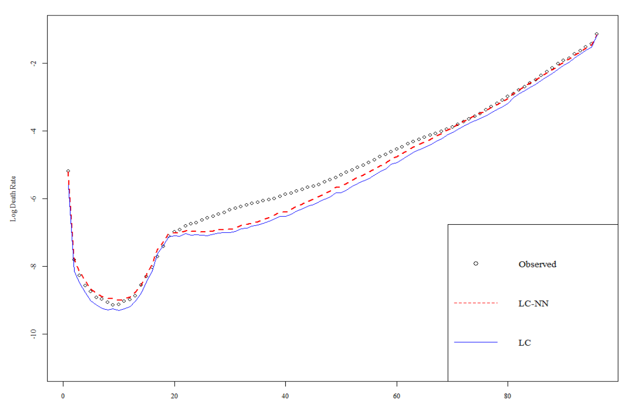


Figure 5: Observed and Forecasted Log of Death Rates for USA in 2021

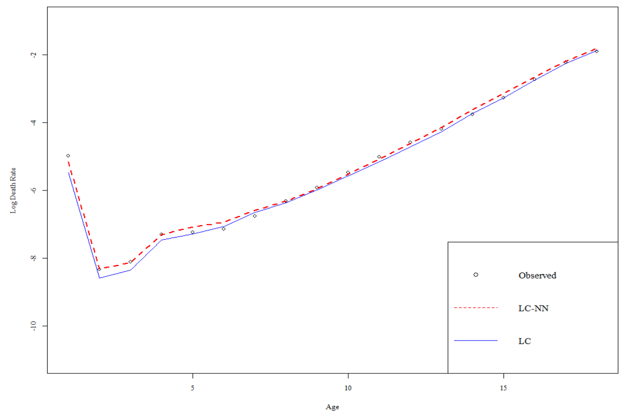


Figure 6: Observed and Forecasted Log of Death Rates for Malaysia in 2015

In short, Figures 2-6 depict the actual and forecasted death rates for different countries and years, demonstrating an ongoing trend of the NN model outperforming the LC model in forecasting accuracy. In each case, both models capture a forecast upward trend in mortality rates with age, but the NN model consistently aligns better with the actual data points. This shows the LC-NN model can identify fluctuations and nonlinearities in the data, resulting in smoother forecasts and a more accurate fit to the actual data. Whether in Germany in 2020, Japan in 2022, South Korea in 2020, the USA in 2021, or Malaysia in 2015, the NN model constantly exceeds the LC model. This indicates that the NN model's ability to analyse complex patterns and correlations in data allows it to produce accurate forecasts of mortality rates across countries and years. Thus, for forecasting mortality rates, the NN model appears to be the most appropriate approach due to its deeper forecasting accuracy and ability to efficiently capture underlying data patterns. Nonetheless, a critical interpretation of model results is required, noting inherent limits and the possible effect of endogenous incidents, such as the continuing COVID-19, on death patterns (Iwamoto, 2023).

3.2. *Performance Evaluation of LC-Model and LC-NN Model*

Table 2 is obtained by data analysis using R-Studio and clarifies the performance study of overall population trends using the LC model and integrated LC-NN model. A thorough examination of the comparison between actual and forecasted death rates is outlined, considering demographic differences across many countries. The forecasting power of the LC and integrated LC-NN model is greatly enhanced by these analytical representations, which show how well it can anticipate demographic changes in the complexity of different national populations.

Table 2: Accuracy measures based on RMSE and MAPE

Country	**GERMAN**		**JAPAN**		**MALAYSIA**	
	LC	LC-NN	LC	LC-NN	LC	LC-NN
RMSE	0.1426	**0.0955**	0.5191	**0.2945**	0.1259	**0.0960**
MAPE	0.0210	**0.0140**	0.0577	**0.0350**	**0.0161**	0.0164

Country	**SOUTH KOREA**		**USA**	
	LC	LC-NN	LC	LC-NN
RMSE	0.1135	**0.1026**	0.1651	**0.1494**
MAPE	0.0119	**0.0107**	0.0217	**0.0206**

*Smallest value are in bold

The RMSE and MAPE values offer a measure of the accuracy of the LC and LC-NN models in forecasting mortality rates for different countries. Germany demonstrates a comparatively low RMSE of 0.1426 in the LC model and an even lower 0.0955 in the LC-NN model, indicating a high level of accuracy for forecasting mortality patterns. The MAPE values for Germany likewise demonstrate this pattern, with the LC model at 0.0210 and the LC-NN model at 0.0140, further confirming enhanced precision with the LC-NN model. Japan exhibits a distinct trend, as seen by an RMSE of 0.5191 in the LC model and a substantially reduced 0.2945 in the LC-NN model. This demonstrates that the LC-NN model significantly enhances the accuracy of predictions compared to the LC model. Japan's MAPE values are 0.0577 for the LC model and 0.035 for the LC-NN model, which confirms the improved accuracy achieved with the LC-NN method.

Malaysia presents a fascinating case, as it exhibits a moderate RMSE of 0.1259 in the LC model and a further improved value of 0.0960 in the LC-NN model. However, the MAPE values show no significant improvement, with the LC model at 0.0161 and the LC-NN model at 0.0164. This implies that the accuracy of the LC-NN model is only slightly impacted. This improvement may be influenced by the way the data is divided or provided in age-grouping forms. The segmentation could have varying effects on the predictive models in comparison to countries that gather data differently.

The RMSE values for South Korea are 0.1135 in the LC model and 0.1026 in the LC-NN model, indicating a moderate level of accuracy. The MAPE values for South Korea are 0.0119 for the LC model and 0.0107 for the LC-NN model. The United States has an RMSE of 0.1651 in the LC model and 0.1494 in the LC-NN model, suggesting a somewhat lower level of accuracy in its forecast compared to Germany and South Korea, but still within acceptable limits. The MAPE values for the United States are 0.0217 for the LC model and 0.0206 for the LC-NN model, indicating a little enhancement with the LC-NN model.

Furthermore, by comparing the RMSE values of the LC and LC-NN models within each country, we may gain insights into the efficacy of LC-NN modelling in improving forecast accuracy. In nations like Germany as well as South Korea, where both models provide low RMSE values, the LC-NN model demonstrates a significantly lower error rate, highlighting its effectiveness in enhancing forecast accuracy. In contrast, in nations such as Malaysia, although the RMSE values are greater, the LC-NN model still performs better than the LC model, although by a small margin.

This paper demonstrates differences in forecasting accuracy among different countries, underscoring the need for customised modelling approaches that consider the distinct demographic and healthcare characteristics of each region. Also, the findings highlight the superior performance of the LC-NN model in improving forecast accuracy, especially in Germany and South Korea, where it outperforms the LC model. These insights show that combining the NN approach with LC coincides with our objectives, as it enables educated, data-driven decision-making processes and improves mortality forecasting methods to effectively address global public health concerns.

4. Conclusion

In the area of mortality studies, numerous hypotheses have emphasised the continuing rise in life expectancy, confirming that advancements in mortality are truly developing. Nevertheless, there is ongoing controversy regarding the pace at which these enhancements happen. This topic is of critical importance as it directly influences our ability to accurately forecast future trends in life expectancy. Therefore, it is essential to focus on the non-linear estimation of the time-dependent parameters within the LC model. Unfortunately, researchers have mostly focused on enhancing the model's accuracy using historical data, disregarding the equally crucial task of forecasting future trends (Nigri et al., 2019). The LC model has been widely used for estimating death rates. This research has significance for examining the complex and extensive consequences of changing mortality patterns, considering the widespread use of the LC technique and other forecasting approaches among actuaries. The integration of NN with the LC model for mortality forecasting has numerous noteworthy implications. This technique improves the model's capacity to detect complex patterns and trends in mortality data that may have been overlooked by the original LC model. Significant progress has been made in mortality modelling to identify patterns in mortality data and provide precise forecasting of mortality rates in the future. Recently, there have been significant changes in forecasting mortality rates. This model uses a time series approach (ARIMA) to clarify changes in age-specific mortality based on an underlying measure of mortality known as the overall mortality index. The research's significant findings are summarised below:

i. The study compares the performance of the LC and LC-NN models in forecasting mortality rates across Germany, Japan, Malaysia, South Korea, and the United States. It consistently demonstrates that the LC-NN model outperforms the traditional LC model in terms of forecast accuracy, as indicated by lower Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) values across most countries.
ii. The research highlights the efficiency of neural network models, specifically a multilayer feed-forward network, in capturing patterns and nonlinearities in mortality data. This capability enables more precise forecasts, particularly in countries with diverse demographic landscapes and varying mortality trends.
iii. By analysing mortality data from multiple countries, the research provides insights into global mortality trends and the efficacy of different forecasting methodologies. This global perspective is crucial for understanding and addressing public health challenges associated with ageing populations, declining birth rates, and other demographic shifts.
iv. The findings have significant policy implications for healthcare planning and resource allocation. Accurate mortality forecasts can aid governments and healthcare providers in anticipating future healthcare needs, developing effective policies, and ensuring sustainable healthcare systems.

Integrating the LC model with a multilayer feed-forward network enhances forecasting accuracy by leveraging the strengths of both methods. However, challenges may arise when addressing the complex temporal dependencies and inherent non-linearities present in mortality data. Recent advancements, such as the integration of recurrent neural networks (RNN) and long short-term memory (LSTM) networks with the LC framework, have demonstrated the efficacy of machine learning in improving long-term mortality forecasts (Marino et al., 2023). Future research could benefit from incorporating LSTM networks into the LC-NN approach for mortality modeling and forecasting, which is expected to significantly enhance the accuracy and robustness of forecasts. It should be noted that while the LC model is widely utilized for mortality forecasting, it is limited by its reliance on prior trend patterns and lack of consideration for external factors such as environmental and health impacts (Hartawan et al., 2023). Including mortality determinants in forecasting future mortality might lead to a more precise model.

Acknowledgment

The authors express their gratitude to the staff of Universiti Teknologi MARA for their support throughout this research. We also acknowledge the Human Mortality Database (HMD) and the Department of Statistics, Malaysia (DOSM), for providing the necessary data.

References

Alho, J. M. (1990). Stochastic methods in population forecasting. *International Journal of Forecasting*, 6(4):521–530.

Baran, S., Gall, J., Ispany, M., and Pap, G. (2007). Forecasting hungarian mortality rates using the lee-carter method. *Acta Oeconomica*, 57(1):21–34.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying lee-carter under conditions of variable mortality decline. *Population studies*, 56(3):325–336.

Deprez, P., Shevchenko, P. V., and Wüthrich, M. V. (2017). Machine learning techniques for mortality modeling. *European Actuarial Journal*, 7:337–352.

Hainaut, D. (2018). A neural network analyzer for mortality forecast. *ASTIN Bulletin*, 48(2):481–508.

Hartawan, I. G. N. Y., Pujawan, I. G. N., Pranata, K. M., and Jayanta, K. (2023). Forecasting population mortality rates using generalized lee-carter model. *Enthusiastic: International Journal of Applied Statistics and Data Science*, pages 16–24.

Hong, W. H., Yap, J. H., Selvachandran, G., Thong, P. H., and Son, L. H. (2021). Forecasting mortality rates using hybrid lee–carter model, artificial neural network and random forest. *Complex & Intelligent Systems*, 7:163–189.

Hyndman, R. and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. OTexts, 3rd edition.

Ibrahim, N. S. M., Lazam, N. M., and Shair, S. N. (2021). Forecasting malaysian mortality rates using the lee-carter model with fitting period variants. In *Journal of Physics: Conference Series*, volume 1988, page 012103. IOP Publishing.

Iwamoto, K. (2023). Deaths jumped 8.9% in japan in 2022 to almost double birth total.

Koissi, M.-C., Shapiro, A. F., and Högnäs, G. (2006). Evaluating and extending the lee-carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics*, 38(1):1–20.

Lee, R. and Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, 38(4):537–549.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting u. s. mortality. *Journal of the American Statistical Association*, 87(419):659–671.

Levantesi, S. and Pizzorusso, V. (2019). Application of machine learning to mortality modeling and forecasting. *Risks*, 7(1).

Marino, M., Levantesi, S., and Nigri, A. (2023). A neural approach to improve the lee-carter mortality density forecasts. *North American Actuarial Journal*, 27(1):148–165.

Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S., and Perla, F. (2019). A deep learning integrated lee-carter model. *Risks*, 7(1):33.

Richman, R. and Wuthrich, M. V. (2019). Lee and carter go machine learning: Recurrent neural networks. *Available at SSRN 3441030*.

Richman, R. and Wüthrich, M. V. (2018). A neural network extension of the lee-carter model to multiple populations. *Annals of Actuarial Science*, 15(2):346–366.

Safitri, L., Mardiyati, S., and Rahim, H. (2018). Forecasting the mortality rates of indonesian population by using neural network. *Journal of Physics: Conference Series*, 974(1):012030.

Torri, T. and Vaupel, J. W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting*, 28(2):519–531.

Yaacob, N. A., Jaber, J. J., Pathmanathan, D., Alwadi, S., and Mohamed, I. (2021). Hybrid of the lee-carter model with maximum overlap discrete wavelet transform filters in forecasting mortality rates. *Mathematics*, 9(18):2295.

Zhang, X., Huang, F., Hui, F. K., and Haberman, S. (2023). Cause-of-death mortality forecasting using adaptive penalized tensor decompositions. *Insurance: Mathematics and Economics*, 111:193–213.

INFLUENCE OF THE KOREAN WAVE ON PURCHASING INTENTIONS OF KOREAN PRODUCTS

Nurul Syahirah Abdullah [1], Nur Alin Yasmin Azam [2] and Noor Aisyah Idris [3,*]
[12,3]College of Computing, Informatics and Mathematics Universiti Teknologi MARA Negeri Sembilan Seremban Campus, 70300 Negeri Sembilan
*noora9995@uitm.edu.my

Abstract

This study aims to identify whether the Korean waves (K-Drama, K-Pop, K-Food, K-Beauty and K-Fashion) and gender influences consumers purchasing intention for Korean products. This study used a quantitative research method to gather quantifiable data from the online survey with a total of 252 respondents who participated in the survey. High quality, beauty, and trend were the top three considered factors when purchasing Korean products. The study also found that Korean products are mostly bought once a month, with a monthly spending of less than RM 50. Multiple linear regression was used to determine the significant influences of the Korean Wave and gender on consumer's purchasing intention for Korean products. The results showed that only K-Drama, K-Beauty and K-Fashion have a significant influence on consumers' buying intention towards Korean products that they have purchased. The study also found no significant difference in students' purchasing intentions for Korean products among the three faculties (Kolej Pengajian Pengkomputeran Informatik dan Matematik(KPPIM), Fakulti Sains Pentadbiran dan Pengajian Polisi(FSPPP), Fakulti Sukan dan Rekreasi(FSR)), suggesting faculty affiliation doesn't impact their purchasing decisions. Therefore, the facilitated digital era plays an important role in spreading Korean content on social media and plays a vital role in influencing consumer's purchasing intention. Researchers suggested that, should use a larger sample size and longitudinal studies to understand the effects of the Korean Wave on youth purchasing behaviors. As a result, this research can gain insight into how the Korean wave influences consumer purchasing patterns and the intention to purchase Korean products.

Keywords: Korean wave, K-Drama, K-Pop, K-Food, K-Beauty, K-Fashion, Purchasing intention

1. Introduction

The Korean Wave, also known as Hallyu in Chinese, has gained global popularity due to its blend of entertainment, culture, and lifestyle. South Korea's cultural representations are increasingly popular among the younger generation, resulting in a growing interest in Korean culture in Malaysia (Zainia et al., 2020). In the late 1990s, during South Korea's financial crisis, the Korean Government introduced the Korean Wave as a soft power tool (Jang and Paik, 2012). It has spread to Southeast Asia regions, including Vietnam, Thailand, Malaysia, and Indonesia (Suh et al., 2006).

In Malaysia, the Korean Wave was introduced by a Korean drama called "Winter Sonata" in 2004. In the mid-2000s, K-pop groups such as TVXQ, Super Junior, and Girls' Generation has also gained dedicated fans. Throughout the past few years, Korean entertainment has greatly influenced cultural shifts. The popularity of the Korean Wave has also influenced consumers' purchasing intention across various Korean product categories (Yu et al., 2012). Korean food (K-Food) has been loved by all Malaysians since its premiere as prime-time television. The fascination with Korean entertainment, including Korean Pop (K-pop) sensations and immersive Korean dramas (K-dramas), has witnessed a notable upswing in Malaysia. Lidwina (2021) also found that Malaysia has the sixth-highest percentage of consumers interested in purchasing Korean products. There are a few tangles of influences influencing Hallyu's relationship with Malaysian consumer intentions. From skincare and cosmetics to fashion and electronics, music, and food, Korean products have found a willing market. Korean content and its promotion are immensely important to spreading Korean content among Malaysian fans, enhancing the Korean Wave's influence and consumers perceptions of Korean products.

Among the goals of the study are to describe consumer patterns of purchasing Korean products, determine the influence of the Korean Wave and gender on consumer purchases of Korean products, and determine if faculty differences affect consumers' purchasing intentions which includes Faculty of Sports Science and Recreation Association (FSR), College of Computing, Informatics and Mathematics (KPPIM) and Faculty of Administrative Science and Policy Studies (FSPPP). Understanding how the Korean Wave affects youth purchasing intention is important. It shows how media and entertainment shape both cultural and economic landscapes by influencing people's tastes and purchasing intention. Due to youths' exposure to new trends and cultural influences, purchasing intention is significantly impacted. Businesses can tailor their marketing strategies by understanding what motivates young people to buy these goods.

2. Literature Review

Originating from South Korea's entertainment industry, Hallyu has evolved into a global force influencing consumer behavior and intentions to purchase Korean products. This literature review aims to examine existing research and hypotheses that highlight the multifaceted influence of the Korean Wave on consumer purchase intentions.Consumer purchase intention measures the desire and tendency to buy a product or service within a specific time frame (Smith, 2023). Several studies have examined how various factors influence purchase intentions, especially in the context of Korean culture and its global impact.

Research by Ing et al. (2018) revealed that the Korean Wave significantly affects attitudes toward Korean products among global consumers familiar with Korean culture. For example, Kwon and Lee (2012) reported that Korean dramas generated $133 million in income in 2010, up from $105 million in 2008. These dramas often feature cultural items and paid product placements, which attract young consumers and encourage them to purchase the featured products (Che Wan Mohd Khair, 2022).Additionally, elements of K-Drama and K-pop, such as idols with fashionable appearances and attractive looks, increase imitation, intention, and positive attitudes toward Korean products (Palasendaram, 2023). By contextualizing and visualizing cultural and entertainment exports with growing global appeal, K-pop significantly impacts the business world (Janiszewski, 1993).

Korean cuisine's popularity in Asia has grown due to international trade, globalization, immigration, and tourism (Verbeke and Poquiviqui Lopez, 2005). The rising demand for convenient home meal replacement products has made Malaysia an emerging market for K-Food (Osman et al., 2014). Platforms like YouTube and TikTok have become crucial channels for promoting K-Food culture (Ingrassia et al., 2022). Korean Beauty (K-Beauty) has also gained popularity due to trends such as the 10-step skincare program, glass skin, and snail mucin Ramaholimihaso. The rising popularity of K-Pop and K-Drama has further boosted the demand for Korean beauty products (Lim et al., 2020). A study by Gerstle (2016) indicates that K-Beauty products are the most popular among Malaysians, driving Korean cosmetic brands into the market and increasing purchase intentions. K-fashion, a South Korean fashion trend that combines traditional and innovative styles, significantly influences Korean fans and is shaped by popular artists. Jobst (2021) found that 54.2% of Malaysians consider Korean fashion popular, with 48.3% favoring Korean-designed items. Gender differences also affect consumer behavior, with distinct views on trust, social influence, and performance expectations influencing purchasing intentions (Hwang and Lee, 2017). Agrippina (2020) found a positive correlation between gender and the intention to purchase Korean music-related products, with women more likely to engage with Korean pop culture. In summary, the study examines how various factors, such as K-Beauty, K-Fashion, K-Drama, K-Pop, K-Food, and gender, influence the independent variable of purchase intention.

3. Methodology

This exploratory research design aimed to uncover insights and generate new ideas about a problem without providing a final solution. A quota sampling technique was employed, focusing on specific strata to ensure the inclusion of subgroups with different characteristics. Respondents were selected from three faculties at University Technology MARA (UiTM) Negeri Sembilan Branch, Seremban Campus: FSR, KPPIM, and FSPPP. The sample consisted of 52.69% FSR students, 26.7% KPPIM students, and 20.6% FSPPP students.

The study aimed to gather data on the purchasing patterns of students at UiTM Negeri Sembilan Branch, Seremban Campus. The target population included 3,161 students from FSR, 1,602 students from KPPIM, and 1,236 students from FSPPP. The sample quota was set at 127 FSR students, 64 KPPIM students, and 49 FSPPP students. The questionnaire for this study was adapted from past research by (Sunhwa, 2021) and (Osman and Ismail, 2022). An online survey was chosen for its cost-effectiveness, speed, accuracy, and flexibility. The research tools included a demographic survey, patterns of respondents purchasing Korean products, and the influence of the Korean Wave on respondents' intention to buy Korean products. Data analysis employed Multiple Linear Regression (MLR) and Kruskal-Wallis test using IBM Statistical Package for the Social Sciences (SPSS) software.

To achieve the first objective, descriptive analysis was used to understand consumer patterns of purchasing Korean products at UiTM Negeri Sembilan Branch, Seremban Campus. Descriptive statistics described the main purposes, frequency of purchases, and monthly spending on Korean products using frequency tables. The second objective involved using MLR to analyze the significant influences of the Korean and demographic factors (gender) on consumer purchasing intentions regarding Korean products. The model assumed a linear relationship between the dependent and independent variables, with a p-value less than 0.05 indicating a significant relationship. The third objective of the study was to examine differences in consumers' purchasing intentions between the faculties (FSR, KPPIM, FSPPP) using the Kruskal-Wallis test, due to the non-normal distribution of the data. This test assessed the influence of faculty exposure to Korean Wave elements and gender on purchasing intentions for Korean products. The Kruskal-Wallis test calculated the sum of ranks for each group, with N as the total number of observations, k as the number of groups, R_i as the sum of ranks for each group, and n_i as the number of observations for each group. The p-value, determined by comparing the test statistic to the chi-square distribution with k-1 degrees of freedom, indicates significance if it is less than 0.05. A significant p-value indicates that at least one faculty group has a different median from the others (Smalheiser, 2017).

$$H = \left(\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right) - 3(N+1) \tag{1}$$

4. Results and Discussion

4.1. *Descriptive Analysis*

Table 1 presents the main reasons for purchasing Korean products, categorized by gender, revealing distinct preferences. High quality is identified as the most significant reason, especially among males, with 29.93% (41 respondents) citing it as their primary motivation. In contrast, beauty is a major consideration for females but significantly less so for males, with only 8.03% (11 respondents) indicating it as a key factor. Both genders consider good packaging, although it does not dominate their decision-making process. Low price is a less compelling factor for females, with only 5% (19 respondents) noting it as a primary consideration, and it is not a significant priority for males. Increasing confidence is relatively more important for females, with 15% (36 respondents) citing it, compared to 10% (11 respondents) of males. Interestingly, trends are the least compelling reason for both genders, showing the lowest percentages overall. These findings suggest that while quality is paramount for all consumers, gender differences are evident in considerations related to beauty, confidence, and price.

Table 1: Main Purposes when Purchasing Korean Products based on Gender

Purpose	Female (%)	Male (%)
Low Price	3.97	2.19
Beauty	28.81	21.90
Good Packaging	13.5	14.60
High Quality	27.97	29.93
Increasing Confidence	18.16	23.36
Trend	7.52	8.03

Table 2 presents the purchasing behavior of consumers with respect to Korean products over various time frames. The data reveals that the highest percentage of respondents, 34.52%, corresponding to 87 individuals, purchase these products once a month. This indicates a significant portion of consumers engage in regular monthly shopping. Conversely, the lowest percentage, 6.35%, consists of respondents who purchase Korean products more than three times a month, highlighting a smaller yet more dedicated group of frequent buyers.

Table 2: Number of Respondents based on the Frequency of Korean Products Purchases

Frequency	Percentage (%)
2-3 times a month	21.83
More than 3 times a month	6.35
Once a month	34.52
Once in a year	16.67
Once in three months	20.63

Table 3(a) illustrates the distribution of respondents based on their spending per transaction on Korean products, while Table 3(b) depicts their average monthly spending on categories such as food, music, cosmetics, and clothing. Analysis of these tables indicates consistent spending patterns among the respondents. Specifically, 85.71% of respondents spend less than RM 250 monthly on Korean products. For per-transaction spending, the majority are in the lower brackets, with 42.46% spending less than RM 50 and 41.27% spending between RM 50 and RM 100. Both tables show a small percentage of high spenders: only 1.19% of respondents spend more than RM 550 monthly, and merely 3.17% spend more than RM 200 per transaction. In summary, the data suggest that most respondents prefer modest spending on Korean products, favoring frequent, smaller purchases over occasional large expenditures. This indicates a market trend towards affordability and regular buying behavior.

Table 3: Number of respondents based on their spending on Korean products

3(a): Based on average spending in a month

Spending (RM)	Percentage (%)
Less than RM 250	85.71%
Between RM 250 - RM 300	6.35%
Between RM 301 - RM 350	5.56%
Between RM 401 - RM 450	0.79%
Between RM 500 - RM 550	0.40%
More than RM 550	1.19%

3(b): Based on spending at a time

Spending (RM)	Percentage (%)
Less than RM 50	3.17%
Between RM 50 - RM 100	9.92%
Between RM 101 - RM 150	41.27%
Between RM 151 - RM 200	42.46%
More than RM 200	3.17%

4.2 *Testing the reliability of the instrument*

On 2 April 2024, the researchers has conducted a pilot study that involved 20 respondents who are students from Universiti Teknologi MARA (UiTM) Negeri Sembilan Branch, Seremban Campus. The results of this study showed that each item in the questionnaire has a Cronbach's Alpha value greater than or equal to 0.7. Table 4.1 shows the result of the reliability test for each variable. Based on the table, it indicates that the consistency of the items in the scale are acceptable.

Table 4: Reliability test

Variable	Cronbach's Alpha value	Reliability
K-Drama	0.915	Reliable
K-Pop	0.975	Reliable
K-Food	0.928	Reliable
K-Beauty	0.942	Reliable
K-Fashion	0.968	Reliable
Purchasing Intention	0.784	Reliable

4.3. *Significant of the Korean Wave and gender on consumer purchasing intention*

The purpose of this study was to determine whether the Korean wave and gender had a significant effect on consumer purchasing intentions. MLR analysis was conducted to confirm these assumptions. Initially, all assumptions were met except for normality and outliers. Thus, removing the outlier was considered to be taken in action to fulfill the assumption. After removing the observation, the data finally satisfied all of the assumptions. To ensure assumptions were met, the process was reanalyzed multiple times.

After all the assumptions are satisfied, MLR were done. Regression analysis produced the following output shown in Table 5. The effects of K-Pop, K-Food, and gender on consumers' purchasing intentions on Korean products are not significant since the p-value exceeds alpha (0.05). In other words, these factors do not statistically influence whether or not people purchase Korean products. Different consumers have different preferences and interests. Not everyone who enjoys K-Drama, K-Beauty or K-Fashion will necessarily be interested in K-Pop or K-Food. This variability can dilute the overall impact of these factors on purchasing intentions. The study also revealed that purchasing intentions for Korean products are not heavily influenced by gender. Hence, marketing strategies might not need to differentiate between male and female consumers as both genders have similar attitudes and behaviors towards these products, leading to gender not being a significant factor.

Table 5: Summary of variable significance

Variable	P-value	Significance
K-Drama	0.018	Yes
K-Pop	0.095	No
K-Food	0.114	No
K-Beauty	< 0.001	Yes
K-Fashion	< 0.001	Yes
Gender	0.323	No

4.4. ***Differences of consumer purchasing intention between faculties***

As part of this study, the researchers also investigated differences in consumer purchasing intentions between students from three faculties: KPPIM, FSPPP, and FSR. The data were carefully prepared and analyzed using Kruskal-Wallis test which ensured independent observations within each group.

The Kruskal-Wallis test was done and obtained a p-value of 0.135, as shown in Table 6, which is greater than the common alpha level of 0.05, indicating that there is no statistically significant difference in consumer purchasing intentions between the students from the three faculties (KPPIM, FSR, and FSPPP). These findings suggest that differences in faculty do not significantly affect student's purchase intentions on Korean products. There is a possibility that students across a range of faculties consume media in a similar manner, leading to a homogenous exposure to Korean products through K-dramas, K-pop, social media, and advertisements. Compared to an older age group with more diverse tastes and preferences, UiTM Seremban students may have more homogeneous tastes as they generally fall within similar age range. It is common for younger consumers to have similar interests and to be influenced by popular culture in similar ways.

Table 6: Krukal-Wallis test

Test statistics	df	P-value
4.010	2	0.135

5. Conclusion and Recommendation

5.1. *Conclusion*

This study examines the influence of the Korean Wave on purchasing intentions for Korean products, focusing on K-Drama, K-Pop, K-Food, K-Beauty, and gender as independent variables. Descriptive statistics indicate that the primary motivators for purchasing are low price, beauty, and increased confidence, with most respondents spending between RM 50 to RM 100 per purchase and maintaining monthly expenditures under RM 250. MLR analysis revealed that while linearity and multicollinearity assumptions were satisfied, initial violations of normality and homoscedasticity were corrected following outlier removal. Nonetheless, the analysis determined that the Korean Wave and gender do not significantly influence purchasing intentions, suggesting no need for gender-specific marketing strategies. Furthermore, Kruskal-Wallis test showed no significant differences in purchasing intentions among students from three different faculties, indicating that faculty affiliation does not impact purchasing intentions for Korean products.

5.2. *Recommendation*

The Korean Wave, encompassing K-drama, K-beauty, K-pop, and other cultural exports, has significantly influenced global consumer choices. To better understand its impact, future research should include larger, balanced sample sizes to ensure more generalizable and accurate results, despite the higher costs and time requirements. Longitudinal studies are recommended to observe changes in youth purchasing behaviors over time, revealing the lasting effects of the Korean Wave. Additionally, methods such as in depth interviews and focus groups can uncover the emotional connections consumers have with Korean products. Researchers should also explore how knowledge of Korean culture and media influences brand preferences and trust, providing valuable insights for effective marketing strategies.

References

Agrippina, Y. R. (2020). Indonesian consumer intention towards south korean pop music. *International Journal of Trend in Research and Development.*, 7(4):106–110.

Che Wan Mohd Khair, W. N. (2022). A systematic review on the influence of product placement (ppl) in korean drama and its effects towards youth purchase intention on korean products in malaysia. In *Forum Komunikasi*, volume 17, pages 87–99. Faculty of Communication and Media Studies (FCMS). Gerstle, T. (2016). Asia personal care and cosmetics market guide. 2016. *US Department of Commerce*.

Hwang, Y. M. and Lee, K. C. (2017). Using eye tracking to explore consumers' visual behavior according to their shopping motivation in mobile environments. *Cyberpsychology, Behavior, and Social Networking*, 20(7):442–447.

Ing, G. P., Azaze @ Azizi, A. A., and Osman, Z. (2018). Korean wave and malaysian young adults: Attitudes, intention and behaviour. *Malaysian Journal of Business and Economics (MJBE)*, 5(1):77– 77.

Ingrassia, M., Bellia, C., Giurdanella, C., Columba, P., and Chironi, S. (2022). Digital influencers, food and tourism—a new model of open innovation for businesses in the ho. re. ca. sector. *Journal of Open Innovation: Technology, Market, and Complexity*, 8(1):50.

Jang, G. and Paik, W. K. (2012). Korean wave as tool for korea's new cultural diplomacy. *Advances in Applied Sociology*, 2(03):196.

Janiszewski, C. (1993). Preattentive mere exposure effects. *Journal of Consumer research*, 20(3):376– 392.

Jobst, N. (2021). South korea: Korean fashion popularity in malaysia 2019.

Kwon, M.-y. and Lee, R. (2012). K-drama leads “hallyu”.

Lidwina, A. (2021). Korean wave dorong orang indonesia beli produk korea selatan: Databoks.

Lim, C.-S., Loo, J.-L., Wong, S.-C., and Hong, K.-T. (2020). Purchase intention of korean beauty products among undergradaute students. *Journal of Management Research*, 12(3):19–40.

Osman, I., Osman, S., Mokhtar, I., Setapa, F., Shukor, S. A. M., and Temyati, Z. (2014). Family food consumption: Desire towards convenient food products. *Procedia - Social and Behavioral Sciences*, 121:223–231. International Halal Conference InHAC 2012.

Osman, S. and Ismail, N. F. M. b. (2022). The influence of korean pop culture on purchasing behavior of korean products among consumers in klang valley. *International Journal of Academic Research in Business and Social Sciences*, 12(10):12–29.

Palasendaram, N. (2023). Letter: Balancing between k-pop and local heritage. Malayasiakini.

Ramaholimihaso, H. (2023). Korean beauty: What is k-beauty? SpaceNK.

Smalheiser, N. R. (2017). Chapter 12 - nonparametric tests. In Smalheiser, N. R., editor, *Data Literacy*, pages 157–167. Academic Press.

Smith, L. (2023). What is buying intent: Definition, data and interpretation. Snov.io.

Suh, C.-S., Cho, Y., and Kwon, S.-H. (2006). The korean wave in southeast asia an analysis of cultural proximity and the globalization of the korean cultural products.

Sunhwa, S. (2021). *The Influence of Korean Wave on Korean Products purchasing intention to Thai consumers in Bangkok*. PhD thesis, Mahidol University.

Verbeke, W. and Poquiviqui Lopez, G. (2005). Ethnic food attitudes and behaviour among belgians and hispanics living in belgium. *British food journal*, 107(11):823–840.

Yu, H., Kim, C., and Kim, H.-R. (2012). Investigating cosmetics purchase of international tourists in south korea. *International journal of management cases*, 14:398–410.

Zainia, F. H. M., Azinuddinb, M., Mior, N. S., Adzim, W. M., and Zaind, W. M. (2020). The role of the hallyu phenomenon in shaping a destination image among malaysian adolescents.

LABOUR FORCE PARTICIPATION RATE AND UNEMPLOYMENT RATE: A MALAYSIAN PERSPECTIVE

Aqila Nawwara Ahmad Fauzi[1], Siti Nur Najiha Razali[2] and Nurul Nisa' Khairol Azmi[3,*]
[1,2,3]College of Computing, Informatics and Mathematics, Teknologi MARA Negeri Sembilan
Seremban Campus, 70300 Negeri Sembilan
*nurulnisa@uitm.edu.my

Abstract

During the COVID-19 outbreak in early 2020, Malaysia's unemployment rate increased from 3.26 per cent in 2019 to 4.54 per cent in 2020. This study aims to forecast the unemployment rate for the next four quarters. Additionally, to determine the impact of the labour force participation rate on the unemployment rate. Data for this study were collected quarterly from the Department of Statistics Malaysia (DOSM) covering the period from 2015 to 2023. The methodologies employed include vector autoregressive (VAR) and impulse response function (IRF) analyses. The findings reveal that the unemployment rate was forecasted for four quarters ahead and the IRF analysis indicates that economic shocks have a significant and lasting impact on the unemployment rate.

Keywords: Vector Autoregressive, Impulse Response Function, unemployment rate, labour force participation rate.

1. Introduction

Despite broad terms in economic development, youth unemployment remains a serious issue in Malaysia. According to Kit (2024), 76 per cent of Malaysians between the ages of 15 and 24 are unemployed, or 10.6 per cent of the country's total unemployment population. An increase in recent graduates joining a declining labour market contributes to this problem. Instability is brought on by high unemployment rates, which have an impact not just on individuals but also on the economy and society across the country, (Rohaimi et al., 2022). In order to predict and tackle these issues, policymakers utilise forecasting models.

The COVID-19 pandemic hit Malaysia hard in March 2020, causing businesses to implement unpaid leave and staff reductions to survive financially. Malaysia's economy shrank by 5.6 per cent in 2020, worse than the 4.4 per cent decline the year before. Moreover, the unemployment has serious social consequences, including increased poverty and family strain. Thus, research is needed to understand how the pandemic has affected the overall employment trends. Government actions and economic stimulus efforts are crucial in helping Malaysia recover and create new job opportunities.

This research tackles two primary objectives. Firstly, it aims to forecast the unemployment rate for the next four quarters. Second, the study delves into how the labour force participation rate influences the unemployment rate. Employers and HR departments will find these insights useful as they offer evidence-based solutions to properly handle future unemployment issues in Malaysia. Hence, this research focuses on examining the relationship between unemployment rate and the labor force participation rate and applied Vector Autoregressive (VAR) and Impulse Response Function (IRF) methodologies using data from the Department of Statistics Malaysia (DOSM).

2. Literature Review

The labour force is an important economic statistic for forecasting future trends, formulating policy, and maintaining competitiveness. By examining significant variables and their relationships, previous research has investigated the relationship between the unemployment rate and labour force participation rate. The percentage of the labour force that is actively looking for work is represented by the unemployment rate, which is an important economic indicator. It acts as an indicator of the state of the economy; large numbers denote trouble in the economy, while low numbers show health (Jamaludin et al., 2021).

According to the Ma'in et al. (2021), persistent difficulties, especially for recent graduates, highlight the need for focused interventions to minimise long-term unemployment risks despite efforts to reduce unemployment through different programmes. The percentage of working-age people who are actively employed or looking for employment in an economy is measured by the labour force participation rate, or LFPR. Future labour market conditions can be predicted with the help of LFPR insights, which also impact policy decisions on labor-related matters (Yusuf et al., 2020).

Aaronson et al. (2006) outlined the percentage of people over 15 to 65 who are not in institutions and who report being either employed or actively seeking for employment is known as the labour participation rate. Ahn and Hamilton (2022) said that the bias in the labour force participation rate has grown over time and that the typical figures understate both the unemployment rate and the rate of participation by an average of almost two percentage points. Malaysia has undergone a significant shift in the composition of its labour force, with an increase in the proportion of female workers. In spite of this development, women's labour market participation rates are still comparatively low, at around 46 per cent (Qinfen, 2017).

In conclusion, the LFPR is a crucial metric that indicates the percentage of working-age individuals who are either employed or actively seeking employment. It provides valuable insights into economic health, labour trends, policy decisions, productivity and social impacts. Understanding the LRPF helps in predicting future labour market shortages or surpluses and supports informed decision-making regarding labour market issues.

3. Methodology

This study aims to explore the relationships among various labor force attributes. Data for this study were sourced from the Department of Statistics Malaysia (DOSM) February 2023 Malaysia Labour Force Statistics Report and quarterly data spanning from 2015 to 2023. The variables covered include unemployment rate and the labor force participation rate.

3.1. *Vector Autoregressive (VAR)*

The linear interdependencies between several time series are captured by the VAR model. It offers a framework for examining dynamic impacts by highlighting the ways in which each variable influences others and itself over time. In a VAR model, each variable has an equation that describes how it has changed over time based on both its own lags and the lags of all other variables and treating them equally. The VAR model describes the evolution of a set of k variables measured over the same sample period, denoted as $VAR(p)$.

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t \tag{1}$$

where,
x is the intercept.
A_i is the square matrix.
ϵ_t are the error terms satisfying the conditions.
c is constant.

The VAR model involves seven steps before forecasts can be made as depicted in Figure 1. The VAR model involves identifying and defining relevant variables, ensuring the time series data is stationary and differencing will be done if the data is nonstationary. Selecting the appropriate lag length (p) based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), another name for BIC is Schwarz Criterion (SC). The model uses the lowest values of AIC and BIC to ensure the best fit.

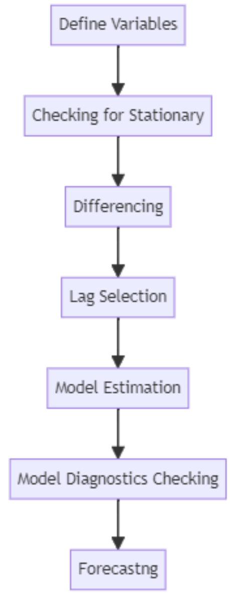


Figure 1: The Flowchart of Steps in VAR

3.2. *Impulse Response Function (IRF)*

The impulse response function (IRF) is a tool used in signal processing, engineering, and economics to analyze how a system responds dynamically to sudden inputs or shocks. It examines how shocks affect variables within the system, aiding in understanding dynamics, forecasting, and evaluating models. By showing the long-term effects of a shock on variables, particularly its impact on unemployment, the IRF helps uncover the dynamics and direction of their influence.

When the shock is observed, the estimation of IRF is defined as:

$$zt = \Phi zt - 1 + ut \quad \text{for } t = \ldots, 0, 1, 2, \ldots, T \tag{2}$$

where Φ is an nxn matrix of coefficients and ut is an $nx1$ vector of reduced form shocks, which is partitioned as

$$ut = r\epsilon t + \zeta t \tag{3}$$

where ϵt denotes the observed shock of interest which is uncorrelated with ζt, where Φ is an nxn matrix of coefficients and ut is an $nx1$ vector of reduced from shocks, which is partitioned as $ut = r\epsilon t + \zeta t$, where ϵt denotes the observed shock of interest which is uncorrelated with ζt (Choi and Chudik, 2019).

4. Result and Discussion

4.1. *Stationary*

In time series analysis, 'stationary' refers to a statistical property of a time series where its mean, variance and autocorrelation is fluctuating around constant number. Figure 2 and Figure 3 depict the result of ACF of the unemployment rate and labour force participation rate respectively. Both figures, which is Figure 2 and Figure 3, show a gradual decline in the ACF, indicating that both time series are non-stationary. Since the time series is found to be non-stationary, differencing technique was used to transform it into a stationary series.

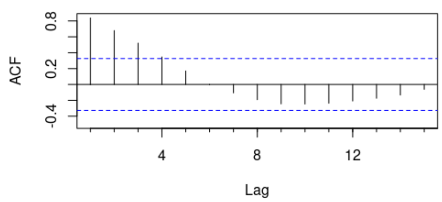


Figure 2: ACF of Unemployment Rate

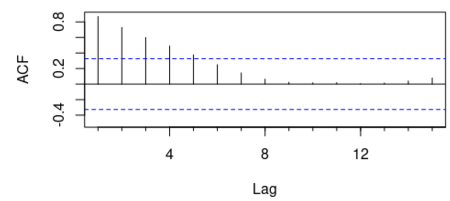


Figure 3: ACF of Labour Force Participation Rate

Figure 4 shows that from 2015 to early 2020, the labor force participation rate and unemployment rate were stable. However, both rates sharply increased starting in early 2020 due to the economic effects of the COVID-19 pandemic. The red line shows a sudden spike in the unemployment rate, and the turquoise line shows significant shifts in the labor force participation rate. These findings highlight the significant impact of the pandemic and economic changes on employment trends and recovery efforts.

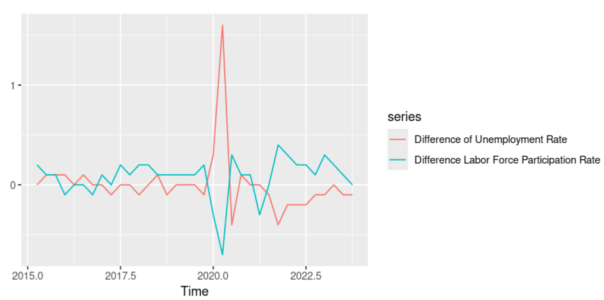


Figure 4: Plot Differencing Labour Force Participation Rate

4.2. *Differencing*

Differencing is a technique used in time series analysis to transform a non-stationary series into a stationary one. Figure 5 and Figure 6 demonstrate that both time series are now stationary as the ACF for both time series show no significant spike. ACF of unemployment rate and labour force participation rate after the first differencing show no significant spike indicate that both of time series are now stationary.

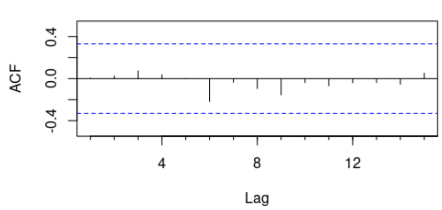


Figure 5: ACF of Unemployment Rate (After First Differencing)

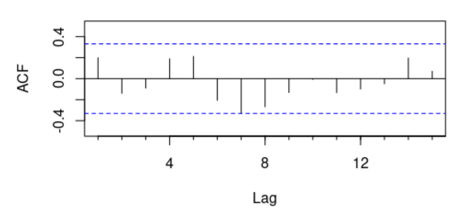


Figure 6: ACF of Labour Force Participation Rate (After First Differencing)

4.3. *Lag Selection*

The value of AIC and SC were observed to determine the lag length. AIC refers to Akaike Information Criterion while SC means Schwarz Criterion. The lower value of both AIC and SC, indicates that the model is better (Fanchette et al., 2020). Table 1 shows that lag 1 has been selected by both AIC and SC.

Table 1: Model Selection Criteria

AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

4.4. *Model Estimation*

Model estimation refers to the process of determining the values of parameters within a statistical or economic model. Model estimation was interpreted in this study to understand the relationship between unemployment rate and labour force participation rate. The output of the model is as depicted in Table 2.

Table 2: Estimation Results for Equation: Unemployment Rate

	Estimate	Std. Error	t value	Pr($>|t|$)
Unemployment Rate	0.848	0.091	9.343	1.17×10^{-10} ***
Labour Force Participation Rate	0.004	0.072	0.056	0.956

From the Table 2, the equation can be written and expanded as follows :

$$\begin{aligned}
\Delta y_t &= 0.287 + 0.848\Delta y_t + 0.004\Delta x_{t-1} \\
y_t - y_{t-1} &= 0.848(y - y_{t-1}) + 0.004(x - x_{t-1}) \\
y_t - 0.848y_t &= y_{t-1} - 0.848y_{t-1} + 0.04(x_t - x_{t-1}) \\
0.152y_t &= 0.152y_{t-1} + 0.004x_t - 0.004x_{t-1} \\
y_t &= y_{t-1} + 0.0026x_t - 0.0026x_{t-1} \qquad (4)
\end{aligned}$$

According to the Equation (4), it can be concluded that unemployment rate for today is based on unemployment rate on 1 period before, adjusted by 0.26% of labour force participation rate.

4.5. *Forecast*

The Table 3 represent the estimated unemployment rates for each quarter of 2024 through the employing of point predictions. Therefore, the estimate for first quarter 2024 is 3.37, and it increases progressively every quarter until it reaches 3.53 in fourth quarter 2024.

Table 3: Forecasting of Unemployment Rate

Year	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2024 Q1	3.37	2.98	3.76	2.77	3.97
2024 Q2	3.43	2.92	3.94	2.65	4.21
2024 Q3	3.48	2.90	4.07	2.59	4.38
2024 Q4	3.53	2.90	4.16	2.56	4.49

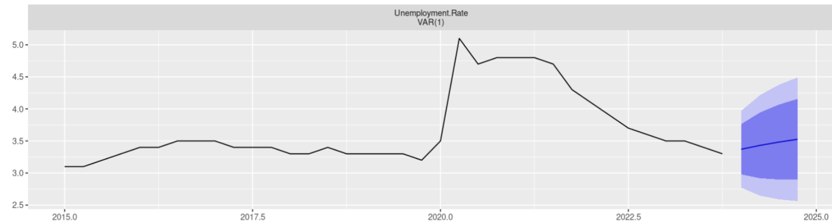


Figure 7: Forecasting Unemployment Rate Based on Labour Force Participation Rate

The shaded regions in Figure 7 show the uncertainty around the point estimates for the confidence intervals. The 80 per cent CI (lighter area) widens from the first to the fourth quarter, indicating increasing forecasting uncertainty. Similarly, the 95 per cent CI (darker area) also expands, highlighting a broader range of possible outcomes as the year progresses. If the labor force participation rate is the main factor considered, the forecast suggests a possible increase in unemployment, with the rate showing a slight upward trend throughout the quarters.

4.6. *Impact towards unemployment rate using Impulse Response Function*

According to the Lu and Xin (2010), the Impulse Response Function (IRF) of the VAR model analyzes the changes in variables after detecting an impulse. The IRF graph determines a variable's behavior in response to a shock by visualizing how the unemployment rate responds to shocks in the labor force participation rate.

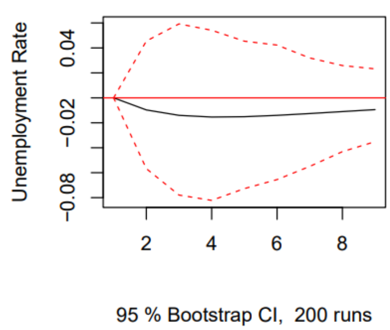


Figure 8: Response of Unemployment Rate to Labour Force Participation Rate Shock

Figure 8 illustrates a significant decrease in the unemployment rate following the initial shock, represented by the black line, which relates to the COVID-19 pandemic. Aaronson et al. (2006) highlighted the importance of understanding labor market dynamics, noting that COVID-19 temporarily reduced the labor force participation rate (LFPR) as individuals exited the workforce for health or family reasons. The black line also shows that the unemployment rate remains below zero for multiple periods, indicating a sustained positive effect. This reflects how the pandemic, company adjustments, government assistance, and economic recovery influenced the labour market. The response is statistically significant and reliable, as the black line stays within the red dashed confidence intervals. This means the effect is not due to chance. Overall, the IRF plot shows a strong and lasting impact of increased job applications during COVID-19 on lowering the unemployment rate.

4.7. *Discussion*

The IRF data shows that an initial shock to the LFPR can briefly increase unemployment as more people enter the labor market, aligning with findings by (Ahn and Hamilton, 2022). The VAR model was used to forecast the unemployment rate for the next four quarters, producing detailed results after ensuring data stationarity. Tables and figures showed the estimated unemployment rates. The IRF analysis also assessed the impact of LFPR, job vacancies, and job applications on unemployment, revealing that shocks to these variables significantly affect the unemployment rate.

5. Conclusions

The purpose of this study was to investigate the connection between Malaysia's unemployment rate and labour force participation. Despite widening confidence intervals indicating long-term uncertainty, the VAR model produced trustworthy insights and stable unemployment projections through 2024. The correctness of the model was demonstrated by its good fit. Economic shocks, particularly during COVID-19, had a major influence on unemployment, according to IRF study. Stable labour markets require constant observation and flexible policy.

6. Recommendations

Advanced analytical techniques are advised in order to comprehend the relationship between Malaysia's unemployment rate and labour force characteristics. Models remain accurate and flexible when methods such as the Bai-Perron test are used to detect weakened structures brought about by changes in the economy or in policy. Additionally, by identifying complex nonlinear correlations between variables, machine learning techniques like Random Forests and Gradient Boosting Machines can increase forecast accuracy. These approaches seek to offer useful perspectives and direct sensible tactics to tackle Malaysia's unemployment issues.

References

Aaronson, S., Fallick, B., Figura, A., Pingle, J. F., and Wascher, W. L. (2006). The recent decline in the labor force participation rate and its implications for potential labor supply. *Brookings papers on economic activity*, 2006(1):69–154.

Ahn, H. J. and Hamilton, J. D. (2022). Measuring labor-force participation and the incidence and duration of unemployment. *Review of Economic Dynamics*, 44:1–32.

Choi, C.-Y. and Chudik, A. (2019). Estimating impulse response functions when the shock series is observed. *Economics Letters*, 180:71–75.

Fanchette, Y., Ramenah, H., Tanougast, C., and Benne, M. (2020). Applying johansen vecm cointegration approach to propose a forecast model of photovoltaic power output plant in reunion island. *AIMS Energy*, 8(2):179–213.

Jamaludin, S., Said, R., Ismail, N. W., and Nor, N. M. (2021). Are jobs available in the market? a perspective from the supply side. *Sustainability*, 13(4):1973.

Kit, E. O. C. (2024). Commentary: Young and jobless in malaysia.

Lu, C. and Xin, Z. (2010). Impulse-response function analysis: An application to macroeconomy data of china. *PhD Thesis*.

Ma'in, M., Othman, S. M., Isa, S. S. M., and Junos, S. (2021). Unemployment and macroeconomic indicators in malaysia. *Global Business & Management Research*, 13(4).

Qinfen, M. (2017). Female labour force participation in malaysia: timeseries evidence. *South East Asia Journal of Contemporary Business, Economics and Law*, 14(3):1–7.

Rohaimi, A., Balqis, A., Nadia, A., and Othman, N. A. (2022). Factors influencing unemployment rate among hospitality industry graduates in malaysia.

Yusuf, M. M., Mohamed, S., and Basah, M. Y. A. (2020). The impact of ageing population on malaysian economic growth. *ASM Sci. J*, 13:1–6.

MONTHLY EXPENDITURE OF PTPTN LOAN RECEIVER AMONG SHAH ALAM UNIVERSITY STUDENTS

Nur Afza Faraheen Afizul[1], Aina Adriana Anizat[2] and Che Norhalila Che Mohamed[3]

[1,2,3]College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus
Seremban, 70300 Seremban, Negeri Sembilan, Malaysia.
*cheno849@uitm.edu.my

Abstract

National Higher Education Fund Corporation, also known as *Perbadanan Tabung Pendidikan Tinggi Nasional* (PTPTN) is one of the study loans taken by students from public and private institutions to finance their cost of living throughout their studies. The specific loan amount has not been revised despite the significant increase in the cost of living, affecting individuals across demographics and causing substantial financial strain. Researchers aim to investigate how students at public and private institutions allocate their overall PTPTN loan. Descriptive statistics were used to describe the total monthly expenditure in nine categories. Public university students exhibited higher variability and maximum expenditure in categories such as food and beverages, savings, transportation, communication, and health. An independent sample t-test showed no significant difference in the average monthly expenditure between public and private university students.

Keywords: Financial behavior, living cost, monthly expenditure, PTPTN.

1. Introduction

The cost of living, which refers to the money needed to pay necessities such as housing, food, taxes, and healthcare, varies greatly depending on location, lifestyle choices, and personal expenditure preferences. This escalating expense has sparked debate and discussion at all levels of society, including among students at higher learning institutions. Students on and off campus face ongoing and increasing financial challenges in the developing world. These difficulties worsened in November 2014, when the Perbadanan Tabung Pendidikan Tinggi Nasional (PTPTN) reduced the maximum loan amounts for degree programs at public and private higher education institutions by 15% and 5%, respectively, except for medical courses at private higher education institutions (Ong et al., 2016) This decline in financial aid has left students struggling with raising costs and limited finances for extra expenditures. To navigate these financial hurdles effectively, students must become adept at managing their finances and using available financial resources responsibly. Financial stress, resulting from these challenges, can significantly impact students' personal lives. For instance, anxiety and worry about their financial situation can lead to sleep problems, where students find it difficult to fall asleep. Furthermore, financial stress can detrimentally affect college students' mental health, contributing to issues such as depression and anxiety (Richardson et al., 2017). Understanding financial stress as a risk factor for poor mental health in college students highlights the negative impact of economic needs on their well-being. Recognizing these effects allows for measures to be taken to help students better manage their finances and reduce related mental health concerns.

2. Research Design

This research utilizes the theoretical framework created by Zulfaris et al. (2020). A descriptive design was used in this study. Descriptive statistics were mainly used to describe features of a population or an existing problem by gathering and analyzing data structures (Awang, 2012). Descriptive statistics were used to describe the total monthly expenditure of public and private university students. An independent t-test was used to investigate whether there is a significant difference in the average total monthly expenditure between public and private university students.

3. Literature Review

The cost of living refers to the expenditure a person or household must incur to meet basic needs such as food, clothes, housing, and other necessities for survival and comfort (Sabri et al., 2018). The cost of living is usually linked to a country's economic growth. As a country's economy grows, it can be challenging for households to preserve their standard of living, and lifestyle changes can cause the cost of living to fluctuate (Evans et al., 2007).

3.1. Demographic factors

3.1.1. Gender

There can be general trends or stereotypes about expenditure habits based on gender, it's essential to recognize that individual preferences and circumstances play a significant role. Males and females want different products, and they are likely to have different ways of tasting and obtaining these which reflects their consumption behaviors (Mitchell and Walsh, 2004). In addition, women, in comparison to men, focus more on the enjoyable aspects of buying and possess a stronger emotional motivation for shopping (Dittmar et al., 2004). It is essential to avoid generalizing expenditure patterns based solely on gender. Individual goals, interests, and financial circumstances significantly influence expenditure habits. Cultural, societal, and personal factors also play a role in shaping these behaviors. 53.5% of female students allocate more money to social activities and leisure while 34.6% of male students do (Sereetrakul, 2013). For instance, a study found that recent technological advances have raised student expenditure. It noted that male students tend to purchase expensive equipment, while female students are more inclined to shoes, bags, and clothing to look presentable for class (Shahryar, 2014).

3.1.2. Age

There is a condition to follow in applying for PTPTN where the age does not exceed 45 years old on the date of application. Students usually have a range of ages between 20-30 years old (Arnaud et al., 2001). It is supported by a study from Omran (2016), 67 respondents, or 69.8% had the greatest age range of 22–23 years old, with 20–21 years old coming in second. There is also a significant relationship between the factors affecting expenditure (age) and respondent background when the p-value for age <0.005 which is equal to 0.021 (Omran, 2016). According to Greenberger and Steinberg (1986), young people often use up money immediately upon receiving it, frequently making unwise financial decisions and wasting resources Hayhoe et al. (2000). Young people are becoming more impulsive shoppers because of having more credit cards, pocket money, or credit cards from other family members (Shim et al., 2010). Despite their intelligence and independence, most young people lack financial literacy. Additionally, university students often use up their discretionary income on personal desires rather than saving for their education.

3.1.3. Types of institution

In Malaysia, there is a notable difference in expenditure between public and private institutions. (Salleh, 2022). For example, UiTM, a public institution, charges just RM590 for six to eight semesters and all courses attended by full-time degree students, while full-time diploma students pay RM540 (Nurul Jannah, 2023). On the other hand, according to the official website of Management and Science University (MSU), the cheapest school fees for a diploma in 2023 were approximately range RM 20000 - RM 25000, while the costliest might reach RM 70000 - RM 75000. Meanwhile, the cheapest educational expenditure for a degree was approximately RM 35000 - RM 40000, while the costliest might go up to RM 295000. For the record, students at public colleges in Malaysia receive a government subsidy covering approximately 90% of tuition fees, leaving them to pay only 10% (Salleh, 2022). This results in significantly lower tuition costs at public institutions compared to market rates in private institutions. Seman and Ahmad (2017) researched to determine the difference in expenditure habits between public and private university students using a 5-Likert scale question. The results as well as the discussion for public students show that the statement "basic needs such as food and phone bills" has the greatest mean of 3.95 and the lowest standard deviation of 1.131, while "shelter" has the lowest mean of 2.95 and the highest standard deviation of 1.239. On the other hand, the tendency to shop has the greatest mean of 3.93 and a standard deviation of 0.944 among private university students. The mean is 2.03 with a standard deviation of 1.291, indicating a fundamental requirement for intake of food.

3.1.4. Level of study

Certain courses and programs have higher costs due to the nature of required materials, equipment, or facilities. For example, Science, Technology, Engineering, and Mathematics (STEM), programs often demand expensive laboratory equipment and technology, leading to increased tuition fees. Additionally, diplomas are typically much cheaper than degrees due to their lower qualification level and shorter duration (Carnevale et al., 2013). UiTM also provides 107,000 residential college spots, with 34,000 reserved for new Diploma and Bachelor students, 52,000 reserved for students who qualify on merit according to UiTM’s placement criteria, and 21,000 reserved for students who apply for residential colleges (UiTM forms committee for student accommodation issues, 2022). At UiTM, most on-campus students are diploma students who typically do not incur significant transportation costs, as the university provides bus services. Degree students, however, who do not meet the college merit for on-campus accommodation often rely on personal cars or e-hailing services, impacting their finances. Some students reduce fuel costs by carpooling.

3.1.5. Financial family category

Income classification in Malaysia can be divided into three categories. As stated by Salleh (2022), the three categories of income levels: are B40 (bottom 40%), M40 (middle 40%), and T20 (top 20%). Each group represents a percentage of the total income households in Malaysia. B40 groups represent households with monthly incomes less than RM3860, M40 groups represent households with monthly incomes between RM3860 and RM8320, and T20 groups represent households with monthly incomes greater than RM8320. As a result of the study's findings, students who face financial difficulties receive financial aid from the government as well as financial resources from their parents to continue their studies at the school (Jamil et al., 2020). This circumstance demonstrates that the high cost of living makes it difficult for students to control their expenditure by continually analyzing the prices of things before purchasing (mean score 4.30) and students only purchase essential requirements (mean score = 3.98).

3.1.6. CGPA

A good performance in academics is very important to certain students who need to pass one semester and go for another semester. CGPA and GPA and students' test results were used to measure the student's performance (Mushtaq and Khan, 2012). Having a great CGPA does not ensure that the students are good at managing their expenditures and can cause stress. Stress can lead to low performance for students (Agolla and Ongori, 2010). If students are unable to cope with stress, their academic performance will suffer. Students who get less sleep or work late at night are more likely to perform poorly (Trockel et al., 2000).

4. Methodology

Primary data are those that are gathered directly from the source and were not acquired by other parties or researchers Salkind (2010) which is the original data collected by the researcher himself for the study. In this study, primary data was used. The goal of primary research was to disclose fresh information that was supported by studies carried out by other researchers and, in the process, to remove any biases held by the researchers (Driscoll, 2011).

The questionnaire was designed as clearly as possible, with a few sections provided. Each section contains several questions that are important to the study. Most importantly, it must have a demographic profile, and the rest of the questions focused on answering the research questions. This research instrument consists of three parts: Part A was the demographic data, Part B was about household monthly expenditure, and Part C consisted of financial behavior, peer influence, and parent's socialization. A cover letter placed on the front page follows the questionnaire for a better understanding of the purpose of the study. Researchers went over the completed questionnaires to ensure that there were no missing data points that could impact the study.

A total of 425 samples were needed for this study using a formula by Umar and Wachiko (2021) based on the number of populations for each of the universities. However, with a 71.7% response rate, only 305 respondents were involved in this study.

Table 1: Number of Population and Sample

University	Population	Sample size calculation	Sample size in the study
Public	55,108	396	261
Private	4,000	29	44
Total	59,108	425	305

Total monthly expenditures were calculated by the total expenditure across various categories. Begin by summing costs for food and beverages, including groceries and dining. Next, include expenditure for room rental, utilities (electricity, water, internet), and transportation (public transit or vehicle-related costs). Also, accounts for communication (phone, internet), personal care products, healthcare, clothing, social activities, recreation, cultural participation, and expenditures related to home decoration, hardware, and rental maintenance.

5. Result and Discussion

5.1. Demographic

From Table 2, most of the respondents were female with a total of 220 (72.13%), and male with a total of 85 (27.87%) respondents participated in this study. The researchers discovered that most of the respondents were at age 21 years old with 154 (50.49%) respondents. Besides, 53 respondents (17.38%) and 45 respondents (14.75%) of them were from 20 to 22 years of age. Otherwise, another 36 (11.80%) were from the 23 years old category. Finally, the lowest number of respondents were from 19, 24, and 25 years of age with 17 (5.57%) respondents. Most of the respondents were from public institutions with a total of 261 (86%) respondents and private institutions with a total of 44 (14%) respondents. According to the respondent, researchers discovered it was obvious that the highest number of the respondents were bachelor's degree holders with a total number of 261 (86%) respondents. In comparison, 44 (14%) of the respondents had a diploma level. Most of the respondents were B40, which was 172 (56.39%) respondents. Besides, 115 (37.70%) of them were M40, and lastly, 18 (5.90%) respondents as T20 in the financial family category. For CGPA, half of the respondents had CGPA from 3.00 to 3.50 with 163 (53%) respondents. Another 121 (40%) of respondents had a CGPA of 3.50 to 4.00. Other than that, the number of respondents whose CGPA was between 2.50 to 2.99 is 21 (7%).

Table 2: Descriptive Statistics of Demographic

Variables		N	Percentage (%)
Gender	Female	220	72.13
	Male	85	27.87
Age	19	4	1.31
	20	53	17.38
	21	154	50.49
	22	45	14.75
	23	36	11.80
	24	9	2.95
	25	4	1.31
Type of Institution	Public	261	85.57
	Private	44	14.43
Level of Education	Diploma	44	14.43
	Bachelor's Degree	261	85.87
Financial Family Category	B40	172	56.39
	M40	115	37.70
	T20	18	5.90
CGPA	2.50-2.99	21	6.89
	3.00-3.49	163	53.44
	3.50-4.00	121	39.67

5.2. Descriptive Analysis for Total Monthly Expenditure Between Public and Private University Students

Table 3 provides a descriptive analysis of the money expenditure of public and private university students across nine categories. University students spend the highest allocation of their PTPTN loan on course fees, books, and stationery per semester. On average, private university students allocate higher amounts to course fees compared to public university students. with RM930.11 and RM863.04, respectively. However, every month, both universities' students spend more on food and beverages. In contrast to course fees, public university students spend RM330.45 per month compared to RM255.32 for private university students. For room rental and utilities, the mean total monthly expenditure for private university students is more than for public university students, at RM 250.86 and RM 229.30 per month, indicating that private university students allocate more than public university students for room rental and utilities. Aside from that, the mean total monthly expenditure on social, recreation, and cultural participation shows private university students allocated more than public university students, at RM 29.77 monthly. This shows that private university students have allocated more money to social life than public university students depending on lifestyle choices.

The analysis of student expenditure between public and private universities reveals distinct expenditure patterns. Public university students incur a wider range of study-related costs with higher maximum amounts and variability, hinting at potential disparities in tuition and fees. The students also spend more expenditure on food, beverages, and personal care, possibly due to different living conditions or lifestyle choices. In contrast, private university students have higher mean total monthly expenditure in savings, transportation, communication, and health, indicating different priorities or access to resources. Despite similar rental and utility costs, private students save more and exhibit greater variability in their expenditures. These differences underscore the diverse financial landscapes students face based on their type of institution.

Table 3: Descriptive Analysis of Money Expenditure of University Students

Money Expenditure (RM)	Public Universiti Students		Private Universiti Students	
	Mean	Std. Dev	Mean	Std. Dev
Course fees, books, and stationery per semester	863.04	313.65	930.11	401.23
Food and beverages per month	330.45	177.45	255.32	147.39
Room rental and utilities per month	229.30	72.40	250.86	69.18
Transportation per month	44.95	63.287	52.73	67.973
Communication per month	53.01	33.33	59.62	46.81
Personal Care per month	87.43	63.25	72.27	43.17
Healthcare per month	20.42	48.93	48.41	52.76
Savings per month	58.18	77.61	109.43	112.79
Cloth and accessories	4.21	16.32	5.00	14.71
Social, recreation, and cultural participation per month	22.92	38.91	29.77	43.49

5.3. Testing the Difference in the Average Total Expenditure Between Public and Private University Students

The independent sample t-test is used to compare the two population means. This test compares the disparities in average total monthly expenditure between private and public university students. Normality and equality of variances are two assumptions that must be met for an independent sample t-test to be valid.

5.3.1. Normality checking

T-tests are a form of parametric procedure that can be employed when the samples are normal, independent, and have equal variance (Kim, 2015). Table 4 shows that the value of unstandardized residuals is assumed to be normal as it does not stray far from the line. There are very few deviations from the straight line. This shows that the data is distributed normally.

Table 4: Kolmogorov-Smirnov

		Statistics	df	Sig.
Unstandardized Residual	Public	0.560	261	0.057
	Private	0.104	44	0.200

Table 4 shows the Kolmogorov-Smirnov test value. The p-value of Kolmogorov-Smirnov for public institutions is 0.057 while private institutions are 0.200. Since the p-value for public institutions and private institutions are greater than 0.05, it fails to reject the null hypothesis. So, the data is considered to follow normal distribution.

5.3.2. Equality of variances

The Levene's Test (F Test) conducted by Snedecor and Cochran (1967) is used to see if two populations' variances are equal. The equality of variances is another assumption that must be met in a parametric test. Levene's test is used to examine this assumption (Kim, 2015). In this case, the variances are not statistically significantly different, as indicated by the p-value greater than 0.05. Since the p-value from Levene's test is 0.231, we fail to reject the null hypothesis of equal variances. Therefore, the assumption of equal variance is fulfilled.

Table 5: Levene's Test for Equality of Variance

F	p-value
1.441	0.231

5.4. Comparison Mean of Total Monthly Expenditure Between Public and Private University Students

Table 6 presents the mean comparison for public and private university students. The total monthly expenditure was calculated from eight categories as detailed in Table 3. The spending of course fees, books, and stationery was excluded since this variable is assessed per semester. With a total of 261 students at a public university, it shows a mean of RM797.67 of total money expenditure and a standard deviation of 234.95. Meanwhile, a private university has a total of 44 students, showing a mean of RM774.56 for total money expenditure and a standard deviation of 214.18. A higher mean score indicates more favourable attitudes (Giroux and Geiss, 2019). Public universities demonstrate higher overall satisfaction in this survey than private universities.

Table 6: Mean Comparison between Public and Private University Students

	Institution	N	Mean	Std. Dev	Std. Error Mean
Total Monthly Expenditure	Public	261	797.67	234.95	14.54
	Private	44	774.56	214.18	32.29

Table 7 presents the result of an independent sample t-test. There was no significant difference in the total monthly for public university students (mean=RM797.67, s 234.95) and private university students (mean = RM774.56, standard deviation = 214.18), conditions; t (0.611), p - value= 0.542.

Table 7: Independent Sample t-test

	t	df	Sig (2-tailed)	Mean difference	Std. Error Difference
Equal variance assumed	0.611	303	0.542	23.11	37.83

Figure 1 depicts the average monthly expenditure of students in Shah Alam. The research reveals no significant difference in monthly expenditure between students from public and private institutions in the area. Notably, the largest portion of their budget, 52.36%, is allocated to educational expenditure. Other major expenditures include food and beverages (19.17%) and rental accommodations and utilities (14.20%). Smaller portions of the budget are dedicated to social, recreational, and cultural activities (1.43%), personal care (5.11%), communication and information (3.24%), health (1.47%), and clothing and shoes (0.25%). This distribution highlights the financial priorities of students, emphasizing educational costs, essential living expenditures, and leisure activities.

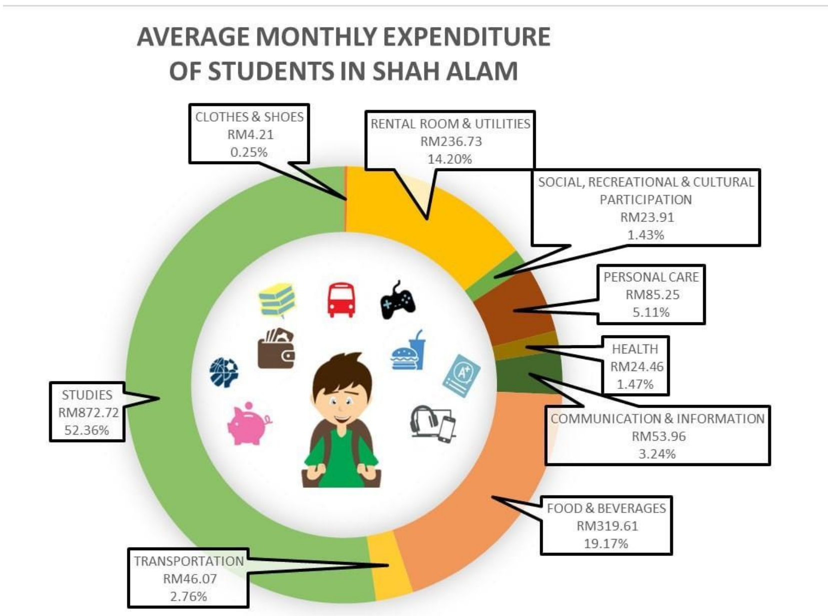


Figure 1: Average Monthly Expenditure of Students in Shah Alam

6. Conclusion

Descriptive statistics were used to achieve this objective. The findings revealed variations in total monthly expenditure between public and private university students across different categories. Public university students exhibited higher variability and maximum expenditure in categories such as food and beverages, savings, transportation, communication, and health. In contrast, private university students typically spent more on studies per semester, with the highest value being RM3070 per semester, compared to RM2800 for public universities. Both groups had similar maximum expenditure on personal care and social, recreation, and cultural participation. Public university students saved more and displayed greater variability in their overall expenditure patterns, highlighting the diverse financial environments between the two types of institutions.

An independent t-test was used to test significant differences in the average total monthly expenditure between public and private institutions. The data for both public and private universities were normally distributed, and all assumptions for the t-test were met. Contrary to what the findings initially suggested, the test results indicated no significant difference in the average total monthly expenditure between public and private university students. This suggests that both groups have similar expenditure patterns, contradicting the earlier observation of higher expenditure among public university students.

7. Recommendations

Suggestions play a crucial role in this research, offering valuable guidance on potential areas for improvement to achieve more accurate outcomes. Therefore, several recommendations have been proposed to address the encountered issues in this study.

While this study utilized resources from Malaysian public and private institutions, there is still room for improvement in terms of scope. Expanding the institutional scope by including more university samples from different institutional contexts and geographical locations can address some of the highlighted shortcomings. Additionally, examining responses from universities in various states, each with its academic setting and sociocultural context, can uncover a wealth of information. A wider range of participating institutions would provide a more comprehensive understanding of students' attitudes and behaviours, thereby enhancing the overall robustness and validity of the research findings.

Acknowledgment

We would like to express our sincere gratitude to everyone who contributed to the success of this study, particularly all the respondents, colleagues, and family, for their unwavering support and encouragement throughout this journey.

References

Agolla, J. E. and Ongori, H. (2010). Assessment of factors influencing employee turnover in public sector organizations: Evidence from kenya. *African Journal of Business Management*, 4(10):2131–2138.

Arnaud, P., Yukawa, Y., Lavie, L., Pélissier, T., Sugiura, M., and Deragon, J.-M. (2001). Analysis of the SINE S1 Pol III promoter from Brassica; impact of methylation and influence of external sequences. *The Plant Journal*, 26(3):295–305.

Awang, Z. (2012). *Research methodology and data analysis second edition*. UiTM Press.

Carnevale, A. P., Rose, S. J., and Hanson, A. R. (2013). Certificates: Gateway to gainful employment and college degrees.

Dittmar, H., Long, K., and Meek, R. (2004). Buying on the internet: Gender differences in on-line and conventional buying motivations. *Sex roles*, 50:423–444.

Driscoll, D. L. (2011). Introduction to primary research: Observations, surveys, and interviews. *Writing spaces: Readings on writing*, *2*(2011):153–174.

Evans, G. W., Brauchle, G., Haq, A., Stecker, R., Wong, K., and Shapiro, E. (2007). Young children's environmental attitudes and behaviors. *Environment and behavior*, *39*(5):635–

658.

Giroux, D. and Geiss, E. (2019). Evaluating a student-led mental health awareness campaign. *Psi Chi Journal of Psychological Research*, *24*(1):61–66.

Greenberger, E. and Steinberg, L. (1986). *When teenagers work: The psychological and social costs of adolescent employment.* Basic Books.

Hayhoe, C. R., Leach, L. J., Turner, P. R., Bruin, M. J., and Lawrence, F. C. (2000). Differences in spending habits and credit use of college students. *Journal of Consumer Affairs*, *34*(1):113–133.

Jamil, N. M., Sulaiman, N., Adznam, S. N. A., & Badari, S. A. Z. (2020). Financial problems associated with food insecurity among public university students in Peninsular Malaysia. *Malays J Nutr*, *26*(3), 411-423.

Kim, T. K. (2015). T test as a parametric statistic. *Korean journal of anesthesiology*, *68*(6), 540-546.

Nurul Jannah Binti Ahmad Jamalullail (2023, November 21). *STRUKTUR YURAN PERDANA*. https://bendahari.uitm.edu.my/index.php/component/content/article/37-pelajar/73-struktur-yuran-perdana?Itemid=101

Mitchell, V. W., & Walsh, G. (2004). Gender differences in German consumer decision-making styles. *Journal of Consumer Behaviour: An International Research Review*, *3*(4), 331-346.

Mushtaq, I., & Khan, S. N. (2012). Factors affecting students' academic performance. *Global journal of management and business research*, *12*(9), 17-22.

Omran, A. (2016). Identifying the main factors affecting the expenditure of students at Universiti Utara Malaysia using PTPTN. *Universitatii Bucuresti. Analele. Seria Stiinte Economice si Administrative*, *10*, 33.

Ong, K. M., Yong, J., Chew, K. Y., & Ng, D. (2016). The sustainability of the PTPTN loan scheme.

Richardson, T., Elliott, P., Roberts, R., & Jansen, M. (2017). A longitudinal study of financial difficulties and mental health in a national sample of British undergraduate students. *Community mental health journal*, *53*, 344-352.

Sabri, B. K. M., Ujang, N., Arif, A. M. M., & Aripin, S. (2018). Compliance constraints among house owners of renovated terrace and semi-detached houses in Selangor, Malaysia. *Journal of Humanities, Language, Culture and Business*, *2*(10), 37-44.

Salkind, N. J. (2010). Content validity. In Salkind, N. J., editor, *Encyclopedia of Research Design*, pages 501–503. SAGE Publications.

Salleh, N. H. M. (2022, July 12). The weight of the subsidy system on public universities. *MalaysiaNow*. https://www.malaysianow.com/news/2022/07/13/the-weight-of-the-subsidy-system-on-public-universities

Seman, A. P. A. and Ahmad, A. B. (2017). Pola perbelanjaan dalam kalangan mahasiswa: Satu kajian perbandingan antara ipta dan ipts. In *Prosiding International Conference on Global Education V*, Universitas Ekasakti, Padang.

Sereetrakul, W., Wongveeravuti, S., & Likitapiwat, T. (2013). Gender differences in saving and spending behaviours of Thai students. *Research in Education*, *90*(1), 68-81.

Shim, S., Barber, B. L., Card, N. A., Xiao, J. J., & Serido, J. (2010). Financial socialization of first-year college students: The roles of parents, work, and education. *Journal of youth and adolescence*, *39*, 1457-1470.

Snedecor, G. W., & Cochran, W. G. (1967). Iowa State University Press. *Ames, Iowa*, 290-327.

Shahryar, S., & Tan, S. T. (2014). Spending behaviour of a case of Asian university students. *Asian Social Science*, *10*(2), 64-69.

Trockel, M. T., Barnes, M. D., & Egget, D. L. (2000). Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviors. *Journal of American college health*, *49*(3), 125-131.

UiTM forms committee for student accommodation issues. (2022, September 20). *The Vibes*. https://www.thevibes.com/articles/education/71920/uitm-forms-committee-for-student-accommodation-issues.

Umar, A. M., & Wachiko, B. (2021). Tara Yamane (1967), Taro Yamane Method For Sample Size Calculation. The Survey Causes Of Mathematics Anxiety Among Secondary School Students In Minna Metropolis. *Mathematical Association Of Nigeria (Man)*, *46*(1), 188.

Zulfaris, M. D., Mustafa, H., Mahussin, N., Alam, M. K., and Daud, Z. M. (2020). Students and money management behavior of a malaysian public university. *The Journal of Asian Finance, Economics and Business*, *7*(3):245–25.

THE DETERMINANTS OF PERSONAL LOAN BORROWINGS: CASE STUDY OF ADULTS IN SHAH ALAM

Jaida Najihah Jamidin[1,*], Aiman Zafira Azizul Rahman[2] and Nur Hannah Mohamad Zamshary[3]

[1,2,3] College of Computing, Informatics and Mathematics Universiti Teknologi MARA, Negeri Sembilan, UiTM Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia

*jaida5698@uitm.edu.my

Abstract

Increasing trends in having personal loans as a way to bridge the gap between income and consumption cost among Malaysian inspired this study. The outcomes can help to enhance credit quality correctly. This study provides survey evidence to identify the determinants of the personal loan borrowings in Shah Alam, Selangor. To find the most reliable predictor of personal loan borrowings, the logistic regression model was used in this study with six main factors considered in this analysis including age, gender, education level, marital status, occupation and income. Online questionnaires were distributed to the 271 respondents in Shah Alam and convenience sampling was the method of sampling used. The collected data were analyzed by using binary logistic regression analysis. The result proved that gender, income, education level and occupation are the significant factors that contributed to the personal loan borrowings contrarily the other two independent variables appear to be non-significant. In addition, the interpretation of odds ratio makes the logistic regression model particularly appealing for modelling. In this study, male are 18.61% times more likely to take out personal loans. Whereas, for one Ringgit decrease in the income of adults in Shah Alam, the odds of person who take out personal loan is multiplied by 0.861.

Keywords: personal loan, Logistic Regression Model

1. Introduction

Personal loans are unsecured loans for personal use that may be used for any reason, whether it be to pay for a wedding, vacation, or consumer durable. Personal loans are highly convenient and can meet all an individual’s needs (Joseph, 2021). A personal loan makes it possible to purchase products and services. Essentially, goods purchased on credit are a debt. Money borrowed from a bank or other financial institution for personal purposes is referred to as a personal loan. Lenders like banks and financial organisations offer personal loans as a sort of consumer lending to help those who are experiencing short-term financial difficulties. It is used to pay for large purchases, family needs, or other significant household costs. These loans might be unsecured or secured by the asset that was acquired, such as collateral, or by a relative who serves as the guarantor (Hashim, 2010). A personal loan is often subject to a flat interest rate, with the rate typically ranking according to the length of the loan or the loan’s funding. The whole amount of interest that has accumulated is due in monthly payments, which must be made until the loan’s term is up. In addition, the loan is issued for domestic, family, or personal usage rather than for commercial or corporate purposes. The broad availability of personal loans is sometimes blamed for the formation of mass markets for consumer goods and the attainment of high standards of living by Western consumers (Beares, 1987).

Malaysians might take advantage of several government benefits, including a first mortgage loan, medical coverage, several incentives and more. Even though they get many benefits, data reveal that household debt has not decreased since Malaysia's household debt to GDP ratio is still high. As of December 2021, Malaysia's household debt as a percentage of GDP was 89%, up from 89.6% in June 2021. In 2020, the Malaysia's household debt-to-GDP ratio reached a new high of 93.2% (Hazim, 2022). This showed that a household utilised around half of its income to pay down debt. As a result, after paying off the debt, there would be less money left over for emergencies, food, transportation, and education. The family would struggle to pay the bills if the working person got sick or lost his job, the risk of loan default increased. In addition, according to research on the ratio of household debt to disposable income, Malaysia was one of the highest levels in the world at 140.4 %, followed by Singapore at 105.3 %, the United States at 123.3%, and Thailand at 52.7 % (Ong, 2010). This shows that each household in Malaysia had loans that are, on average, 1.4 times larger than their annual income (Ong, 2010).

Recent research has found that the key reason for the rise in personal bankruptcies is the inability to repay loans (Mien & Said, 2018). This was because people often found it tough when it was time to make their monthly loan payments due to the high interest rates. On the other hand, some of the contributing causes resulted from inadequate financial planning. The danger of financial crisis and economic instability might then increase the challenges in developing countries during the global financial crisis. Hence, investigating the factors that influence personal loans in Malaysia was important since this study might help peoples to avoid getting into debt problem in the future.

2. Methodology

2.1. Description of Data

The data used in this study was collected from a survey of 271 respondents in Shah Alam. The dataset includes variables such as gender (categorical, male or female), age (continuous, years), and income (continuous, RM). Educational level was coded as "diploma level," "bachelor's degree level," "master's degree level," and "Doctor of Philosophy (PHD) level. In the meantime, marital status was classified as "single," "married," "separated," or "widowed." For personal loan borrowings, 1 if the person takes the personal loan and 0 if the person does not take the personal loan. Besides, the study was conducted from July 2022 to September 2022. The study focused on the Malaysian population aged 18 and above.

2.2. Research Framework

The purpose of theoretical framework is to present the relationships that propose in the study based on the researcher's speculation. It describes a summary of theory on a particular problem that was developed through a review of variables involved. As part of this study, there are one dichotomous dependent variable involved which is the personal loans borrowing and eight independent variable that includes gender, age, education level, marital status, occupation and household income. The theoretical framework of this study is shown in Figure 2.1.

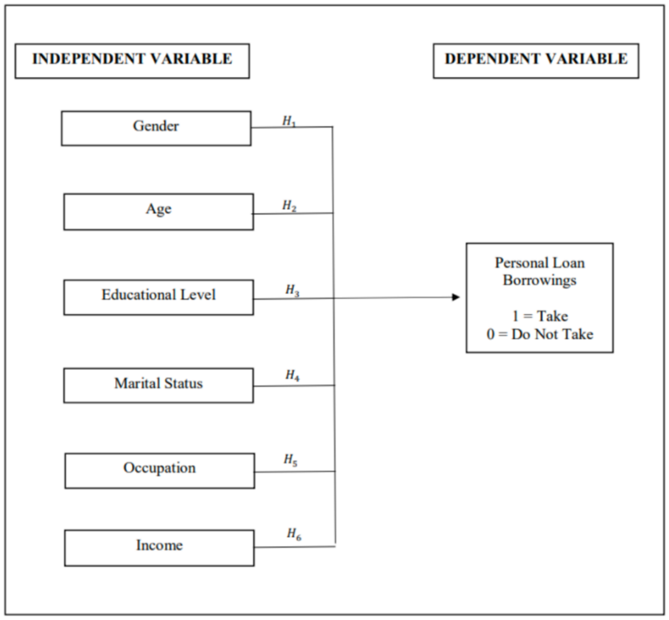


Figure 1: Research Framework

2.3. Logistic Regression Analysis

Logistic regression analysis is one of the regression analyses for the binary classification, which includes two class values for the dependent variable. This analysis' main goal is to define variables and identify the relationship between the dependent binary variable and independent variables. In addition, logistic regression analysis is not the same as linear regression, in that logistic regression produces a constant output, whereas linear regression produces a continuous output. The output of logistic regression only contains a limited number of possible values, and it can only predict between two possible outcomes.

The general form of logistic regression model is:

$$logit\ (P) = \log\left[\frac{P(X)}{1 - P(X)}\right]$$

$$= B_0 + B_1X_1 + B_2X_2 + \cdots + B_kX_k + \varepsilon$$

From the general model, it can be simplified into a preceding model which 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 success dependent variable

B_0= constant of the equation
$B_0, B_1, \dots, B_k$= the coefficient of independent variable
X_1= Age
X_2= Gender
X_3= Education Level
X_4= Marital Status
X_5= Occupation
X_6 = Income

For the model's evaluation, five criteria are considered. The logistic regression model is evaluated using these five criteria which are the Omnibus Test, Hosmer and Lemeshow Test, Predictive Efficiency Model, Wald Statistics, and lastly Cox and Snell R^2 and Nagelkerke R^2.

The Omnibus Test was used to assess the overall model accuracy. The researcher applies the Omnibus Test to examine if any regression coefficients are significantly different from all other coefficients, apart from the coefficient B_0. The null hypothesis (H_0) in this study was that all the regression coefficients were equal to zero, whereas the alternative hypothesis (H_1) was that they were not equal to zero. The null hypothesis (H_0) is rejected and the alternative hypothesis (H_1) is accepted if the p-value for the Omnibus Test is less than 0.05. This revealed that data from independent variables could aid in more precise forecasting of the factor of indebtedness.

Hosmer and Lemeshow Test can be used to check the goodness of fit, in which the researchers intended to see if the data was a good fit test for the logit model or not. The null hypothesis (H_0) was that the logistic regression model matched the data well, whereas the alternative hypothesis (H_1) was that the logistic regression model suited the data poorly.

Classification table compared the logistic regression model's predicted value for the dependent variable with the data set's actual observed value. If the aggregate percentage is 60%, it is a good indicator of predictive efficiency. In addition to the overall percentage, sensitivity, specificity, and accuracy can be used to evaluate the predictive efficiency model. Sensitivity refers to the likelihood that the test will correctly identify personal loan borrowings, or the likelihood that any given set of personal loans will be identified by the test (refer to '1'). While specificity refers to the likelihood that the test will correctly identify someone who is debt-free (refer to "0"),

Wald Statistics displays the statistical significance of each independent, regressor, or predictor variable. Each regression coefficient is evaluated independently by the Wald Statistics, with the null hypothesis (H_0) being a regression coefficient equal to zero and the alternative hypothesis (H_1) being a regression coefficient that is not equal to zero.

Cox and Snell R^2 and Nagelkerke R^2 method displays the amount of variation in the anticipated variable that the model can account for. The range of values is $0 < R^2 < 1$. Depending on the situation, different R^2 values will be considered "good." Even a relatively low R^2 of 0.5 may be seen as quite powerful in some domains, such as the social sciences.

2.4. Odds Ratio

Odds ratio referred as an analysis that measures the relationship between an input and an outcome. It reflects the probability that an outcome will occur given a certain input, as opposed to the odds that the outcome would occur in the absence of that input. The predicted increase in the likelihood of the outcome per unit increase in the value of the exposure is known as the regression coefficient in a logistic regression. In other words, the odds ratio for a one-unit increase in input is determined by the exponential function of the regression coefficient.

The conversion from probability to odds is call as monotonic transformation, which means that as the probability increases, so do the odds. The probability scale ranging from 0 to 1. In this study, the value of '1' means the people are taking personal loans while value of '0' refers to people who are not taking personal loans. Therefore, the probability P(X) can be determined as follows:

$$P(X) = \frac{Number\ of\ outcome}{Number\ of\ all\ outcome}, where\ 0 < P < 1$$

The odds ratio shows that an event is likely to occur in comparison to not occurring. As a result, probability is used to represent the likelihood of something happening or not happening.

$$P(X) = \frac{Probability\ of\ People\ Taking\ Personal\ Loans}{Probability\ of\ People\ Not\ Taking\ Personal\ Loans} = \frac{P(X)}{1 - P(X)}$$

Therefore, the odds ratio shows that the change in odds depends on the outcome when the values of a predictor increased by one unit.

3. Result and Discussion

3.1. Criteria for Model Evaluation

For the model's evaluation, Omnibus Test, Hosmer and Lemeshow Test, Predictive Efficiency Model, Wald Statistics and Cox and Snell $R^2$and Nagelkerke $R^2$are considered. The logistic regression model is evaluated using these five criteria.

Table 1: The Results of Criteria for Model Evaluation

The Criteria for Model Evaluation	Full Model
Omnibus Test	0.00
Hosmer and Lemeshow Test	0.886
Predictive Efficiency Model (%)	69.7
Cox and Snell R^2	0.15
Nagelkerke R^2	0.205

Based on Table 1, the Omnibus Test, Hosmer and Lemeshow Test, Predictive Efficiency Model and Cox and Snell R^2 and Nagelkerke R^2 results show that the full model met all evaluation model criteria. According to the Omnibus Test results, it is possible to predict the determinant of personal loan borrowings more accurately using data from independent variables. As the overall percentage for the complete model is more than 60%, it demonstrates an excellent predictive efficiency, according to Predictive Efficiency Model percentage. Also, the Hosmer and Lemeshow Test expresses that the model adequately described the data because the p-value is above the 5% level of significance. Furthermore, for Cox and Snell R^2Nagelkerke R^2, both R^2 values indicate that all the variance in personal loan borrowing, which is between 15% and 20.5%, can be explained from all independent variables.

3.2. Fitting Logistic Regression Model

Logistic Regression Model (LRM) estimates the probability of an event occurring, which in this study the probability of someone's taking personal loan, based on the independent variables which are Age, Gender, Education Level, Marital Status, Occupation and Income. As for the indicator, the researcher used Female for Gender, SPM for Education Level, Widowed for Marital Status and Employed for Occupation. Table 2 shows full model coefficient which includes the estimated coefficients and the p-values.

Table 2: Full Model Coefficient

Variable	**Estimated Coefficient**	**P-Value**
Age	0.031	0.153
Gender (Male)	0.621	0.038
Education Level (Diploma)	-0.193	0.763
Education Level (Bachelor's degree)	-0.91	0.122
Education Level (Masters' Degree)	0.282	0.671
Education Level (PHD)	0.154	0.836
Marital Status (Single)	1.751	0.169
Marital Status (Married)	2.279	0.061
Marital Status (Separated)	2.148	0.129
Occupation (Self-Employed)	1.499	0.126
Occupation (Retired)	-0.486	0.635
Income	-0.149	0.04
Constant	-3.31	0.111

According to Table 2, the variables marital status and age are not statistically significant in explaining the determinants of personal loan borrowings because their significant 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 four variables of gender, education level, occupation, and income are the main factors that contribute significantly to personal loan borrowings. The logistic model equation used to predict personal loan borrowings is shown below.

Z= - 3.31 + 0.031 Age + 0.621 Gender (Male) - 0.193 Education Level (Diploma) - 0.91 Education Level (Bachelor's Degree) - 0.282 Education Level (Master's Degree) - 0.154 Education Level (PHD) + 1.751 Marital Status (Single)+ 2.279 Marital Status (Married) + 2.148 Marital Status (Separated) + 1.499 Occupation (Self-Employed) - 0.486 Occupation (Retired) - 0.149 Income.

3.3. Interpreting Odds Ratio

The odds ratio is another way to describe the likelihood that people will take out a personal loan. The odds ratio for the full model is shown in Table 4.9.

Table 3: Full Model Odds Ratio

Variable	Odds Ratio
Age	1.032
Gender(Male)	1.861
Education Level (Diploma)	0.824
Education Level (Bachelor's Degree)	0.403
Education Level (Masters' Degree)	1.326
Education Level (PHD)	0.167
Marital Status (Single)	5.763
Marital Status (Married)	9.769
Marital Status (Separated)	8.571
Occupation (Self-Employed)	4.479
Occupation (Retired)	0.615
Income	0.861

The output for the odds ratio shows that married people were 9 times more likely to take out personal loans. When compared to the widowed, divorced people are 8.571 times more likely to take out personal loans. Meanwhile, it seems single people are 5.763 times more likely to take out personal loans compared to widows. Self-employed people are 4.479 times more likely than employed people to take out personal loans. Furthermore, an older person in Shah Alam, a male, and a master's degree holder had a one-time higher odds ratio. Since the odds ratio is higher than 1.0, married individuals, divorced individuals, single individuals, self-employed individuals, age, being male, and being a master's degree holder may be major factors in personal loan borrowing.

4. Conclusion

Knowing what factors affect personal loans is important since many people with loans seldom have access to large disadvantages. Having a personal loan is related with a higher chance of increased debt load, additional monthly payment, potential credit harm, and higher payments than credit cards. This disadvantage extends beyond the financial world to lifestyle and bankruptcy (Esty, 1997). Therefore, taking out a personal loan may have an impact on bankruptcy legislation. Even while financing is surprisingly widespread across all income levels and occupations, education level is still a key consideration when applying for a personal loan. The last factor that may affect personal loan borrowing is marital status.

The study's findings demonstrate that its two goals were adequately attained. This study first determined the key elements that influence the borrowing of personal loans. In order to expose the contributing causes, this study also analysed the relative odd of the incidence of personal loan borrowing.

Based on the findings in the study, some recommendations would benefit for further research. Firstly, it is recommended to increase the sample size so that the findings will have better representative of the population. Since the number of samples used in this study is small, so the likelihood of encountering significant value on which the study is based is low. Secondly, future studies thus encourage that more variables should be added, or different factors must be used in variation that impact personal loan borrowings. Since this study more focus on the socioeconomic factors, hence it is recommended another study be done using new methods and variables. This study was intended to serve as a foundation for subsequent research into personal loan borrowings issue.

As this study has limitations on online surveys, future researcher needs to consider to extent to face-to-face for more reliable study. As a result, it should be explored in future study using face-to-face surveys to address the causes of personal loan borrowings, with a focus on economic reasons, society, and nation. To produce useful information in research, the future researcher should also generate accurate results in the analysis and be aware of the possibility of bias. To assure the validity and accuracy of the study findings, more accurate data analysis results are required.

References

Beares, P. (1987). Consumer lending. American Institute of Banking, American Bankers Association.

Hashim, M. (2010). Personal finance. UPENA.

Hazim, A. (2022, Apr). Retrieved from https://themalaysianreserve.com/2022/04/07/malaysias-debt-to-gdp-ratio-could-stabilise-at-65-5-this-year/

Joseph, J. (2021). A Study on Personal Loan at Bajaj Finserv Limited. *International Journal of Innovative Research in Technology*, 8(6): 442-451

Mien, T. S., & Said, R. (2018). A cross-sectional household analysis of household consumption patterns: An indirect approach to identify the possible factors of personal bankruptcy. *Jurnal Ekonomi Malaysia*, 52(3), 231–246

Ong, T. K. (2010). Household debt: Blame the cars, not a home. The edge.

THE INFLUENCE FACTORS OF SOCIAL MEDIA USERS ON INTENTION TO ADOPT A ZERO WASTE LIFESTYLE IN PETALING, SELANGOR.

Siti Aishah Mohd Shafie[1, *], Aniz Nazirah Borhan[2] and Nur Balqis Basyirah Abdullah[3]

[1,2,3] Kolej Pengajian Pengkomputeran, Informatik dan Matematik,
Universiti Teknologi MARA Cawangan Negeri Sembilan Kampus Seremban
*ctaishah@uitm.edu.my

Abstract

Zero-waste should be practiced in managing waste disposal or production, preventing pollution, conserving natural resources, improving quality of life and well-being. Encouraging a sustainable practice in social media is one of the solutions to raise awareness of a zero-waste lifestyle. In Malaysia, Selangor has been identified as the highest waste-producing state, and within the state, the highest populated district is the Petaling district. As a result, the Petaling district is where this study is being explored. In addition, this study emphasizes gender, age group, peers' impressions of zero-waste posts on social media, and the ability to distribute zero-waste information on social media. This study aims to investigate the factors that influence social media users' intentions to adopt a zero-waste lifestyle in Petaling. In this study, 210 observations were chosen as samples. The method that is being used in this study is logistic regression analysis. Overall, based on the results, it appears that peers' impressions of zero-waste posts on social media serve as a significant influencing element.

Keywords: Intention, Lifestyle, Peer's Impression, Social Media, Zero-waste

1. Introduction

Minimizing waste to zero may seem impossible, especially in light of today's society, which is easily affected by a wide range of consumer goods and service offerings. The worldwide volume of solid waste is estimated to be over billions of tonnes per year (Song et.al, 2015). The term "zero-waste" can be defined as a waste prevention movement that encourages individuals to reuse all items (Jusoh et al., 2020). In general, zero-waste should be practiced managing waste disposal or production, prevent pollution, conserve natural resources, and improve quality of life and well-being (Zulkilfi et al., 2015). Zero Waste Malaysia reveal that Malaysia generates over 38 thousand tonnes of trash every day. The country is unable to cope with the growing amount of solid waste produced each year with the limited number of landfills. It disrupts the environment's ecosystem, which has a detrimental influence on human life, flora and fauna, and animals.

One strategy that can help Malaysia's waste problem is the adoption of a zero-waste lifestyle. In truth, it is possible to produce new goods from a variety of waste materials without endangering natural resources or raising production costs. There are several approaches to promote sustainable practices, including through social media advertising. According to Boran (2018), social media is an important tool for fostering social change. Additionally, Petaling's population is growing every day. As a result of there being more customers, this raises the overall amount of garbage. The pattern of disposal household solid waste in Petaling is the sole topic of this study. Therefore, it is crucial to research how social media affects the adoption of a zero-waste lifestyle.

Zero-waste lifestyle adoption targets to reduce waste, increase recycling, avoid over consumption, and prefer reusable products. However, in order to achieve a zero-waste lifestyle, spreading awareness is crucial, especially on social media. Nowadays, social media scrolling is a habit in people's daily lives. According to survey report by Department of Statistics Malaysia in 2022, it stated in year 2021 the most popular online activity for 99% of Malaysians was using social media. Therefore, social media can increase public awareness of zero-waste initiatives. In this study, some factors that influence social media users on intention to adopt a zero-waste lifestyle were discussed in terms of gender, age group, and peers' impressions on zero-waste post on social media, and ability to distribute zero-waste information on social media.

Frequency is the rate at which something occurs over a specific time while preference means greater likin for one alternative over another alternative. According to a study by Rapada et.al (2021) concerning the influence of social media on consumer behaviour towards plastic pollution, 76% of the 213 respondents that took part in the study used social media daily. The preference of social media is caused by the new opportunities provided by social media itself to connect people with little relation but similar interest (Young, 2021). Morgan and Gartshore (2021) stated that Instagram is the most preferred platform for individuals to increase health awareness compared to Facebook and YouTube. The related posting and information about the zero-waste lifestyle could be reached by users based on their frequency and preference of social media if and only if there are such postings about it.

Attitude was defined as a settled act or opinion about something (Pustaka, 2017). A person's attitude can be classified as either positive or negative. This is because, in order to form behaviour, and change to occur, each individual must have a different attitude or perception of certain knowledge. Awang et.al (2021) state that attitudes have a substantial role in pro-environmental behaviour. Nowadays, social media has a big impact on millennial perceptions. Therefore, social media has played an essential role in improving millennial attitudes toward a zero-waste lifestyle. Changing a person's attitude about zero-waste information through social media enhance their desire to adopt a zero-waste lifestyle (Young, 2021).

Gender is a visible difference between males and females viewed from their values and behaviour. One of the most important considerations in adopting a zero-waste lifestyle is gender. It has contributed significantly to raising environmental awareness. According to survey report by Department of Statistics Malaysia in 2022, individuals' use of information and communications technology (ICT) in 2021 show that males use the internet 97.2% more than females (96.3%). In addition, social media may provide basic knowledge on gender issues and specific environmental protection measures, and it has been discovered that females were more concerned about environmental protection than males (Dhenge et al., 2022). Most of the research shows that females are more likely to adopt a zero-waste lifestyle. Nevertheless, both males and females were highly interested in adopting zero-waste to conserve resources and minimize pollution (Bagagiolo et al., 2022).

There are four categories of human age; child, teenager, adult, and senior adult (Nithyasri and Kulanthaivel, 2012). A child can be classified between the day of birth until 12 years old, a teenager between the ages of 13 until 18 years old, an adult between the ages of 19 until 59 years old, and senior adult between the ages of 60 years old and above. As reported by Department of Statistics Malaysia in 2022, all age groups exceed 90% of using the internet except for senior adults. This shows how prevalent people are on social media in the current times. Social media has played an important part in influencing the attitude of young generations to become more sustainable. Attitudes toward adopting a zero-waste lifestyle must be taught at a young age so that their minds will become more responsive to developing a positive environmental attitude (Dhenge et al., 2022). Thus, they may be educated themselves to adopt a zero-waste lifestyle. Besides, people under 50 years were 75.7% more interested in adopting zero-waste to conserve resources and minimize pollution than people that more than 50 years with only 24.3% (Bagagiolo et al., 2022).

The next aspect is the influence on individual's behaviors towards zero-waste adoption on social media that can affect the individual's peers desire them to act on social media. For example, peers may ask individuals to like, share, or ta them on zero-waste social media postings (Young, 2021). There was a positive significant influence on the peers' impressions on social media towards zero-waste lifestyle regarding recycling (Pratiwi et al., 2021). Furthermore, the perception of behaviour that peers need to act out or participate in environmental practice had a positive impact towards the influence on sustainable waste management (Muniandy et al., 2021). Bagagiolo et al. (2022) agreed that individuals that obtained a recommendation from their peers on social media were significantly impacted by a high desire to adopt a zero-waste lifestyle. According to Young (2021), the ability to distribute zero-waste information on social media is significantly correlated to adopt a zero-waste lifestyle. The ones who have the ability to distribute zero-waste information on social media are likely to apply zero-waste lifestyle in their daily lives, such as eliminating single-use items, recycling, repurposing household items, composting, connection with the zero-waste community and educating others about the lifestyle. The information that is shared by friends or acquaintances on social media is considered as valuable and trustworthy advice that can influence others behaviour to follow the lifestyle (Lahath et al., 2021).

2. Material and Methods

2.1. Description of Data

This study focuses on the gender, age group, peers' impressions on zero-waste post on social media, ability to distribute zero-waste information on social media, and the intentions to adopt a zero-waste lifestyle in the Petaling district as to finding the factors that influence social media users' intentions to adopt a zero-waste lifestyle in the Petaling district. A cross-sectional study design is used in this study; thus the conclusion of a cross-sectional study is only valid while the study is being conducted. The independent variables for this study are gender, age group, peers' impressions on zero-waste post on social media, and ability to distribute zero-waste information on social media while the dependent variable is intention to adopt a zero-waste lifestyle.

The population of this study was the citizens of Petaling district that were aged 18 and above. In the 2020 national census by Jabatan Perangkaan Malaysia in reported by year 2022, it was concluded that the Petaling district had the most citizens in Selangor, hence it is the best district to meet this study's objectives. The sample size for this study is 200 and it is decided by referring to the research done by Sanusi et al. (2022), 200 are used as a sample where its population is all Petaling citizens who are age between 18 and 65. The researcher manage to get 210 Petaling residents as samples of the study and it's fulfilled the minimum sample size required that being calculated. A convenience sampling method was sampling method used in this study to select the respondents.

Primary data was used in obtaining the information to accomplish this study. This study was conducted by distributing the questionnaire via Google Forms to the respondents. This study adopted the questionnaire by Young (2021) to get better understanding on how social media strategies influence people to adopt zero-waste behaviours. Each of the survey questions was established to fulfil the specific research objectives. The Cronbach's Alpha is used in determining the reliability of perceptions of peers' impression on zero-waste post on social media and ability to distribute zero-waste information on social media. The value may range anywhere from 0 to 1, however, only a value that is larger or equal to 0.7 is considered to be reliable (Taber, 2018). The summated score is used in this study. The summated score is applied for independent variables in this research which are peer's impressions on zero-waste post on social media and ability to distribute zero-waste information on social media.

2.2. Theoretical Framework

Figure 1 presents the theoretical framework in this research. The independent variables are gender, age group, peer's impressions on zero-waste post on social media, and ability to distribute zero-waste information on social media. Moreover, the dependent variable is the intention to adopt a zero-waste lifestyle.

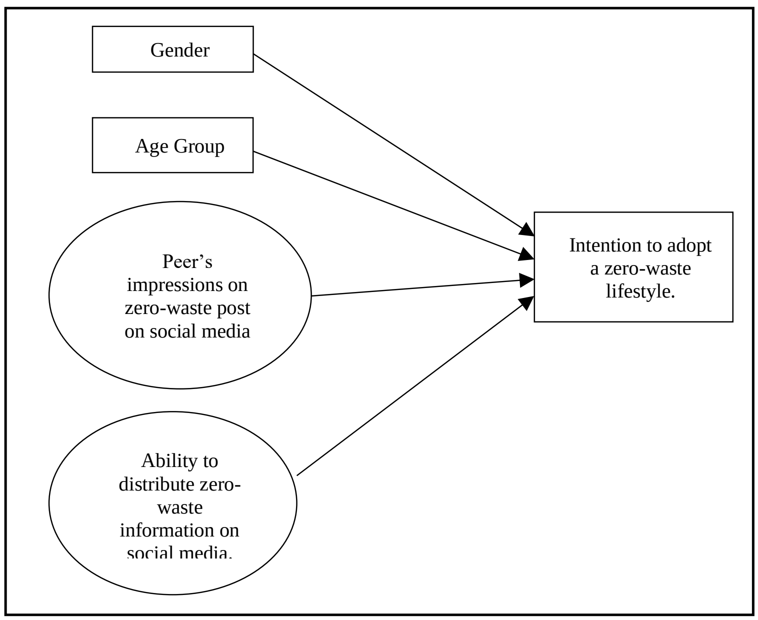


Figure 1: Theoretical Framework (Young, 2021)

2.3. Logistic Regression Model

It is well known that predictive analysis included logistic regression. Regression modelling is appropriate for describing data and elucidating the link between a single binary dependent variable and a number of independent variables (Hidayat, 2017). In this study, logistics regression is carried out using the forward selection technique. First off, the model won't contain any independent variable. The candidate to enter the equation is the predictor with minimum p-value. The independent variable is included in the model of the p-value is less than 0.05. The programme terminates if none of the independent variables is deemed significant enough to be included in the logistic regression equation, which provides the theoretical foundation for this study.

The model evaluation process includes a number of tests that are used to gauge the model's effectiveness and goodness-of-fit (Atmathew, 2015). The significance level for each test was set at 5%. The Omnibus test for model coefficient is the first test in the model evaluation process. The information from the study's explanatory variables gives a stronger prediction of the predicted variable if the p-value for the Omnibus test is smaller than alpha value (0.05) and vice versa. The Hosmer-Lemeshow test is the second test used to gauge how well logistics regression models fit data. The model has a good fit if the p-value of Hosmer-Lemeshow test is greater than alpha value (0.05) (Glen, 2020). In addition, predictive efficiency model used as classification table in deciding which model is better than the other.

The Wald statistics test is to determine whether one or more independent variables in logistics regression are significant. The null hypothesis is rejected and the independent factors are significant in determining the intentions to adopt a zero-waste lifestyle if the Wald statistics test's p-value is less than 0.05. If the odd ratio (OR) is equal 1, it means that the exposure had no effect on the likelihood that the social media user would decide to lead a zero-waste lifestyle. If the OR is more than 1, it means that the exposure is more likely to increase a social media user's intention to live a zero-waste lifestyle, whereas an OR of less than 1 means that the exposure was less likely to increase intention. Last but not least, Cox and Snell R Square and Nagelkerke R Square show how much variance in the predicted variable is explained by the model, with a range of 0 to 1.

3. Results and Discussion

3.1. Descriptive Analysis

There were 210 residents of Petaling who responded, as displayed in Table 1. Male respondents make up 96(45.7%) and female respondents make up 114 (54.3%). According to the proportion, there were significantly more female respondents than male respondents. The age group of 18 to 29 years old has the highest percentage of respondents (in terms of age grouping) with 121 respondents (57.6% of the total). Following this are the respondents between the ages of 30 and 49, who made up 54 of the respondents (25.7%) and the minority of the respondents, who made up 35 of the respondents (16.7%), are over the age of 50. Since the questionnaire were filled out online, the majority of the respondents are between the ages of 18 and 29.

Next, the majority of the respondents are holding a Degree as their highest level of education with 132 respondents (62.9%), followed by Diploma holders with 42 respondents (20.0%). UPSR/ PT3/ SPM are ranked as the top three highest levels of education in this study with the frequency of 17 respondents (8.1%). About 13 respondents (6.2%) are coming from Master holders and the minority of the respondents are coming from other certificates and PhD holders with four and two respondents for each variable, respectively. On the other hand, Shah Alam residents comprise the majority of the respondents, with 134 respondents (63.8%), followed by Petaling Jaya residents with 60 respondents (28.6%). The minority of the respondents are coming from Subang Jaya with a frequency of 16 respondents (7.6%).

Table 1: Demographic Profile of Respondents

Variable	Respondents, n	Percentage, %
Gender		
Male	96	45.7
Female	114	54.3
Age		
18-29	121	57.6
30-49	54	25.7
50 and above	35	16.7
Highest Level of Education		
UPSR/ PT3/ SPM	17	8.1
Diploma	42	20.0
Degree	132	62.9
Master	13	6.2
PhD	2	1.0
Others	4	1.9
City of Residence		
Petaling Jaya	60	28.6
Shah Alam	134	63.8
Subang Jaya	16	7.6

3.2. Reliability Test

The study's reliability test is displayed in Table 2. It demonstrates that every variable has a Cronbach's Alpha coefficient greater than or equal 0.7, indicating a high level of dependability. Thus, it can be said that all the data in this study are reliable and consistent.

Table 2: Cronbach's Alpha Coefficient

Variable	No of Items	Cronbach's Alpha
Peers' impression on zero-waste post on social media	4	0.923
Ability to distribute zero-waste information on social media	4	0.865

3.3. Logistic Regression Analysis Results

The results of logistic regression analysis are displayed in Table 3. The approach entered demonstrated that none of the independent variables, including gender, age group, peers' impression on zero-waste post on social media, and ability to distribute zero-waste information on social media are statistically significant. This is because the significant value for Wald test is greater than 0.05. The intention to adopt a zero-waste lifestyle in Petaling is therefore unaffected by the factors of gender, age group, peers' impression on zero-waste post on social media, and ability to distribute zero-waste information on social media. Because none of the variables are significant, the odds ratio of enter technique will not determine anything.

The forward approach of logistic regression's model reveals that only peers' impression on zero-waste post on social media are statistically relevant in terms of the intention to adopt a zero-waste lifestyle. By contrasting the model, Omnibus test is used to assess the performance of the model. The results of the Omnibus test of enter and forward approach are displayed in

Table 3. Both approaches produce results of 0.019 and 0.006 respectively, which both demonstrate a significant model with a significance value less than 005. The forward approach is also more important that the enter method. As a result, the forward technique will provide a superior model forecast.

The result of the Hosmer-Lemeshow test comparing enter and forward procedure are displayed in Table 3. Enter method's significant value is 0.116, which is greater than 0.05. This indicate that the model and the data are well matched. Additionally, because the significant value for the forward method is less than 0.001, the model does not well match the data.

The results of enter and forward method's Cox and Snell R Square and Nagelkerke R Square are displayed in Table 3. Cox and Snell R Square values for enter and forward procedures are 0.062 and 0.035, respectively. Next, enter and forward techniques' Nagelkerke R Square values are 0.448 and 0.252 respectively. According to the results of the Nagelkerke R Square of the enter method, the independent variables of gender, age group, peers' impression on zero-waste post on social media, and ability to distribute zero-waste information on social media explained 44.8% of the variance related to the intention to adopt a zero-waste lifestyle. In addition, using Nagelkerke R Square of forward technique, it can be concluded that around 25.2% of the variation associated with the intention to adopt a zero-waste lifestyle was explained by peers' impression on zero-waste post on social media.

Table 3 demonstrates that the Wald test's significant value is less than the alpha value, which is 0.05. It can be concluded that peers' impression on zero-waste post on social media is influenced by the intentions to adopt a zero-waste lifestyle among social media users in Petaling. Furthermore, the odds of having intentions to adopt a zero-waste lifestyle are multiplied by 1.562. With one unit increase in peers' impression on zero-waste post on social media, the probability of the intention to adopt zero-waste lifestyles increases by 44.6%.

Table 3: Logistic Regression Analysis Results

Variable	Method					
	Enter			Forward		
	B	Sig.	Exp (B)	B	Sig.	Exp(B)
Gender	17.589	0.996	43541899.70			
Age group						
18-29	-0.050	0.972	0.952			
30-49	16.892	0.997	21673543.49			
Peers' impression on zero-waste post on social media	0.360	0.195	1.434	0.446	0.011	1.562
Ability to distribute zero-waste information on social media	0.022	0.941	1.022			
Constant	-1.998	0.444	0.136	-1.549	0.428	0.213
Goodness of Fit Test	**Enter**			**Forward**		
	χ^2		Sig.	χ^2		Sig.
Omnibus Test	13.058		0.019	7.5000		0.006
Hosmer and Lemeshow	12.875		0.116	27.157		<0.001
	Enter			**Forward**		
Cox and Snell R square	0.062			0.035		
Nagelkerke R Square	0.448			0.252		

4. Conclusion

The primary goal of this study is to discover the elements that affect Petaling social media users' intention to live a zero-waste lifestyle. To do this, logistics regression analysis was performed. The outcome of this aims has demonstrated that since p-value of 0.011, peers' impression on zero-waste post on social media are significant This finding was corroborated by the earlier research by Young (2021), who found that one the key factors influencing intentions to adopt zero-waste lifestyle is peers' impression of zero-waste Instagram posts.

Acknowledgment

The researcher would like to thank to all of respondents who participated in our research. This research would not be complete this study without the help from them.

References

Awang, S., Annuar, S. N. S., & Gisip, I. A. (2021). The effect of social media influencer attributes towards pro-environmental intention. *Rom. J. Inf. Technol. Autom. Control*, 31(1), 111–124.

Bagagiolo, G., Vigoroso, L., Pampuro, N., & Cavallo, E. (2022). The role of social interaction and personal characteristics in affecting the adoption of compost from organic fraction of municipal solid waste in Italy. *Agronomy*, 12(2), 445.

Bersama kita Zero Waste. Zero Waste Malaysia, (2019, Jun). https://zerowastemalaysia.org/bersamakitazerowaste/ Accessed December 2023

Dhenge, S., Ghadge, S., Ahire, M., Gorantiwar, S., & Shinde, M. (2022). Gender attitude towards environmental protection: a comparative survey during covid-19 lockdown situation. *Environment, Development and Sustainability*, 1–46.

Atmathew, A. (2015, Aug). Evaluating logistic regression models: R-bloggers. https://www.r-bloggers.com/2015/08/evaluating-logistic -regression-models/ Accessed January 2023

Hosmer-lemeshow test: Definition. Glen, S. (2020, Dec). https://www.statisticshowto.com/hosmer-lemeshow-test/ Accessed January 2023

Jusoh, Z. M., Osman, S., Arif, A. M. M., & Salleh, R. M. (2020). Household solid waste disposal pattern in Malaysia. *Malaysian Journal of Human Ecology (Mjhe)* Volume 1, 2020, 62.

Lahath, A., Omar, N. A., Ali, M. H., Tseng, M.-L., & Yazid, Z. (2021). Exploring food waste during the covid-19 pandemic among Malaysian consumers: The effect of social media, neuroticism, and impulse buying on food waste. *Sustainable Production and Consumption*, *28*, 519–531.

Morgan, E., & Gartshore, L. (2021). Are they ready? Awareness and preparedness of the emerging workforce for dental check by one. *British Dental Journal*, 1–6.

Muniandy, G., Anuar, M. M., Foster, B., Saputra, J., Johansyah, M. D., Khoa, T. T., & Ahmed, Z. U. (2021). Determinants of sustainable waste management behavior of Malaysian academics. Sustainability, 13(8), 4424.

Nithyashri, Jayaraman, & Kulanthaivel, G., (2012).Classification of human age based on Neural Networking using FG-NET Aging database and Wavelets. 4th International Conference on Advanced Computing,ICoAC2012.1-5.10.1109/ICoAC.6416855

Poket stats negeri selangor. Jabatan Perangkaan Malaysia, D. (2022, May). https://cloud.stats.gov.my/index.php/s/OOL5GS8iBN7puEu#pdfviewer/ Accessed December 2022

Pratiwi, P. Y., Handra, T., & Choirisa, S. F. (2021). Determinants of zero waste lifestyle adoption among generation-z. In Conference series (Vol. 3, pp. 371–384).

Rapada, M. Z., Yu, D. E., & Yu, K. D. (2021). Do social media posts influence consumption behavior towards plastic pollution? *Sustainability*, *13*(22), 12334.

Regresi logistik. Hidayat, A. (2017, Sep). https://www.statistikian .com/2015/02/regresi-logistik.html/ Accessed January 2023

Sanusi, S. F. A., Shafie, F. A., Yatim, S. R. M., Rasam, A. R. A., Kek, T. L., & Leh, O. L. H. (2022, Jun). The Association of Urban Lifestyle with Respiratory Illness in Petaling District. *Malaysian Journal of Medicine and Health Sciences*, 132-144.

Sikap. Pustaka, D. B. d. (2017).https://prpm.dbp.gov.my/cari1?keyword=sikap/ Accessed December 2022

Social media viewed as essential tool for social change. The Irish Times. Boran, M. (2018, Sep). https://www.irishtimes.com/business/ technology/social-media-viewed-as-essential-tool-for-social -change-1.3642006/ Accessed Dec 2022

Song, Q., Li, J., & Zeng, X. (2015). Minimizing the increasing solid waste through zero waste strategy. *Journal of Cleaner Production*, 104, 199–210.

Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education, *Research in Science Education*.48, 1723-1296.

Young, L. (2021). Instagram's impact on intention to adopt zero-waste behaviors (Unpublished doctoral dissertation). University of Hawai'i at Manoa.

Zulkifli, N., Hashim, A. H., Badari, S. A. Z., & Jubadi,W. (2015). Penerimaan dan tingkah laku pengguna terhadap amalan sisa sifar ke arah penggunaan lestari.19th MACFEA National Seminar, 274-279

TIME SERIES FORECASTING OF ROAD ACCIDENT IN MALAYSIA BY USING BOX-JENKINS AND UNIVARIATE MODEL

Nur Farah Bahiah Mohd Zarmi[1], Nur Balqish Zainal[2] and Norliana Mohd Lip[3,*]
[1,2,3]Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Negeri Sembilan, Seremban Branch, 70300 Seremban, Negeri Sembilan.
[*]norliana7287@uitm.edu.my

Abstract

This study aims to predict future road accidents in Malaysia using time series forecasting techniques. The objectives are to determine the best model between Box-Jenkins and Univariate methods, and to forecast road accident numbers based on the model with the lowest error measures of MAPE and RMSE. The study analyzes Malaysia's historical and current traffic accident trends to provide insights for forecasters, government agencies, and transportation firms. Researchers can also build upon the findings to further investigate traffic accident prediction in Malaysia. The results show that the Double Exponential Smoothing model provides the most accurate forecasts for 2024 road accident numbers in Malaysia. This information can help authorities implement targeted interventions to improve road safety and reduce the significant human and economic toll of traffic accidents.

Keywords: Box-Jenkins, ARIMA, Univariate Model, Single Exponential Smoothing (SES), Double Exponential Smoothing (DES)

1. Introduction

A road accident is an unexpected event causing loss or injury without the injured person's fault, for which legal relief may be sought. According to the World Health Organization (2023), approximately 1.3 million people die in road accidents annually. Gimino G. and T.A.(2023) mentioned that Malaysia reported 545,588 road accidents in 2022, resulting in 6,080 deaths, as shared by Transport Minister Anthony Loke. Human error, such as driving under the influence, speeding, and distracted driving, is frequently cited as the main cause of traffic accidents. Unfavorable weather and poor road conditions also significantly impact accident rates.

Malaysian roads experience various types of accidents, each presenting distinct risks. Multi-vehicle collisions from tailgating and inadequate following distances cause chaotic pileups. Head-on collisions, the most critical, result in severe injuries or fatalities, especially at high speeds. Low-speed accidents, though seemingly less severe, still risk casualties, particularly for pedestrians and cyclists. Merging accidents highlight the importance of proper speed and blind spot checks in congested areas. Recognizing and addressing these accident dynamics is crucial for enhancing road safety in Malaysia (Ali, 2023).

Predicting traffic accidents through time series analysis involves identifying patterns and trends in past data. Understanding seasonal, monthly, or daily fluctuations in accident frequency is crucial for developing prediction models. Time series analysis helps forecast peak accident periods, allowing authorities to allocate resources more efficiently during high-risk times. Additionally, it evaluates the long-term effectiveness of safety initiatives, enabling the optimization of road safety strategies. Predictive models provide insights into the temporal dynamics of traffic accidents, leading to focused and timely precautions that improve traffic safety.

This study compares Box-Jenkins and Univariate models to determine the most accurate method for forecasting road accidents. The objectives are to identify the most effective model and use it to forecast future road accidents in Malaysia.

2. Methodology

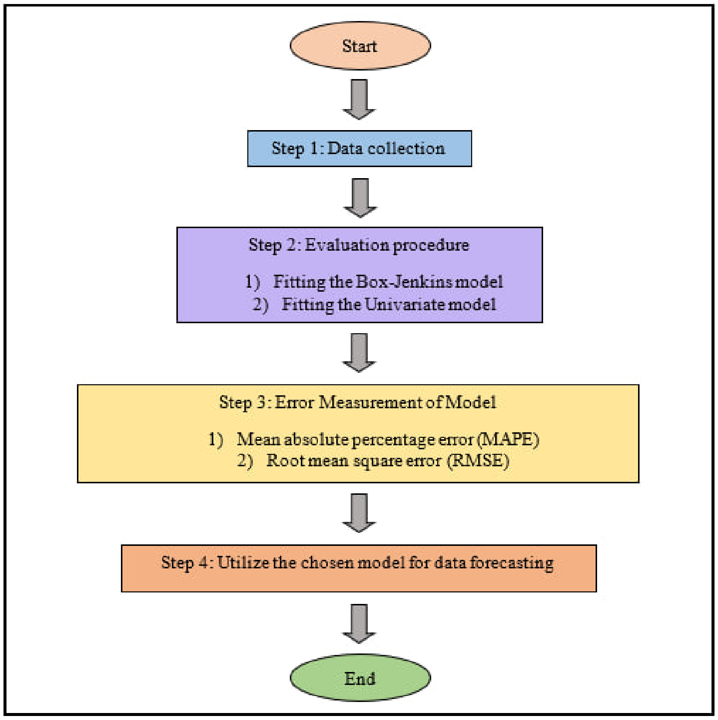


Figure 1: Research Framework Of The Study

2.1. *Data collection*

The data used in this study comes from secondary sources, specifically the Traffic Investigation and Enforcement Department (JSPT) of the Royal Malaysia Police (RMP). The analysis includes monthly data from January 2019 to December 2023.

2.2. *Box-Jenkins Model*

The Box-Jenkins model, also known as the ARIMA (AutoRegressive Integrated Moving Average) model, is a powerful and widely used approach for time series forecasting. Developed by George Box and Gwilym Jenkins, it combines autoregressive (AR) processes, where future values are regressed on their own lagged values; differencing (I), which involves subtracting previous values to achieve stationarity; and moving averages (MA), which model the error term as a linear combination of past errors. The ARIMA model is identified and estimated through a systematic process of model selection, parameter estimation, and diagnostic checking, making it highly adaptable to various types of time series data. Its strength lies in its ability to handle non-stationary data by transforming it into a stationary series, making it suitable for a wide range of forecasting applications. The ARIMA model is expressed by Lazim (2011):

$$Y_t = Y_{t-1} + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{1}$$

where;

Y_t	: the time series at time t,
$\phi_1, \ldots, \phi_p$	: the autoregressive parameters,
$\theta_1, \ldots, \theta_q$	: the moving average parameters,
ϵ_t	: the error term at time t,
$Y_{t-1}, \ldots, Y_{t-p}$	: the lagged values of the time series, and
$\epsilon_{t-1}, \ldots, \epsilon_{t-q}$	: the lagged values of the error term.

The Autoregressive Integrated Moving Average (ARIMA) model, represented as ARIMA (p, d, q), is developed for non-stationary data to achieve stationarity by differencing the variable d times. In this model, p denotes the autoregressive process, and q represents the moving average process. For example, ARIMA(1,1,1) is expressed as:

$$w_t = \mu + \phi_1 w_{t-1} - \theta_1 \epsilon_{t-1} + \epsilon_t \tag{2}$$

where;

$w_t = y_t - y_{t-1}$	: the first differenced series,
μ	: the constant term,
ϕ_1	: the impact of the previous differenced value,
θ_1	: the influence of the previous error, and
ϵ_t	: the current error term.

2.3. ***Univariate Model***

2.3.1. *Single Exponential Smoothing*

The time series forecasting technique known as Single Exponential Smoothing (SES) is applied to univariate data that without a trend or seasonal pattern.The method works by applying exponentially decreasing weights to past observations, which means more recent observations are given more weight than older ones. The smoothing constant, α is the only parameter required. The method can be written as equation (3) mentioned by Rosyid et al. (2019):

$$F_{t+m} = \alpha y_t + (1-\alpha) F_t \tag{3}$$

where;

F_{t+m}	: Forecast value for m period ahead
α	: Smoothing constant
y_t	: Actual value

2.3.2. *Double Exponential Smoothing*

Double exponential smoothing is an extension of single exponential smoothing that is used for time series data with a trend. In order to overcome the drawbacks of single exponential smoothing, a technique to take trends in the data is included. It provides a way to forecast future values by adjusting for both the level and the trend in the data. There are two smoothing constant, α and β that is the parameter required. In order to get the prediction value, step on finding stationary value and trend value, the method can be written as equation (4) and equation (5) mentioned by Rosyid et al.(2019):

Stationary value

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{4}$$

Trend value

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta) T_{t-1} \tag{5}$$

Forecasting value

$$F_{t+m} = S_t + T_t m \tag{6}$$

where;
S_t : Stationary value
T_t : Trend value
y_t : Actual value
α : Smoothing constant
β : Smoothing constant
F_{t+m} : Forecast value for m period ahead

2.4. *Error Measurement of Model*

The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), which are provided in equations (6) and (7), respectively, were used to evaluate the accuracy of the models in this study.

2.4.1. *Root Mean Square Error (RMSE)*

$$RMSE = \sqrt{\frac{1}{n}\sum_{t}^{n}(x_t - \hat{x_t})^2} \tag{7}$$

2.4.2. *Mean Absolute Percentage Error (MAPE)*

$$MAPE = \frac{1}{n}\sum_{t}^{n}|(\frac{x_t - \hat{x}_t}{x_t}) \times 100| \tag{8}$$

where;
x_t : Actual observed value
$\hat{x}_t$: Predicted value
n : The number of predicted values

3. Result and Discussion

The first findings of the study demonstrate the summary of error test that was utilised to determine which ARIMA models were most effective in predicting the frequency of traffic accidents for Box-Jenkins approach. Table 1 illustrates that, when compared to other models, ARIMA (1, 1, 1) was chosen to be the most effective model to use because its RMSE and MAPE values were the lowest.

Table 1: Selected ARIMA Model

Model	RMSE	MAPE
ARIMA(1,1,1)	16239.8200	45.0952
ARIMA(1,1,2)	16528.1000	45.8986
ARIMA(1,1,3)	16812.3500	46.6681

The data in Table 2 shows the RMSE and MAPE for SES and DES models. The DES model outperforms the SES model with lower RMSE (475.9685) and MAPE (1.1338), indicating its superiority for Univariate Models. The optimal α and β values for DES are 0.9596 and 0.0536, respectively. The RMSE value reflects the average error magnitude, while the MAPE value indicates the average percentage error. Despite potential absolute error variations, the DES model maintains accuracy, balancing relative accuracy and absolute error. This demonstrates the effectiveness of the DES model in forecasting time series data.

Table 2: Comparison of SES and DES

Model	RMSE	MAPE
Single Exponential Smoothing (SES)	4065.5544	9.6934
Double Exponential Smoothing (DES)	475.9685	1.1338

Table 3 shows DES has a smaller MAPE and RMSE than ARIMA(1,1,1), with values of 475.9685 and 1.1338, respectively. High accuracy of the model may be attained when the RMSE and MAPE values are at their lowest. As a result, the DES model is determined to be the most accurate model for estimating the number of Malaysia's road accident for 2024.

Table 3: Comparison of ARIMA(1,1,1) and DES

Model	RMSE	MAPE
ARIMA(1,1,1)	16239.8200	45.0952
Double Exponential Smoothing (DES)	475.9685	1.1338

As shown in Table 3 it is obvious that the DES model is the most accurate model for estimating the number of Malaysia's road accident in 2024. The predicted value's outcome is shown in the Figure 2 and Table 4 below.

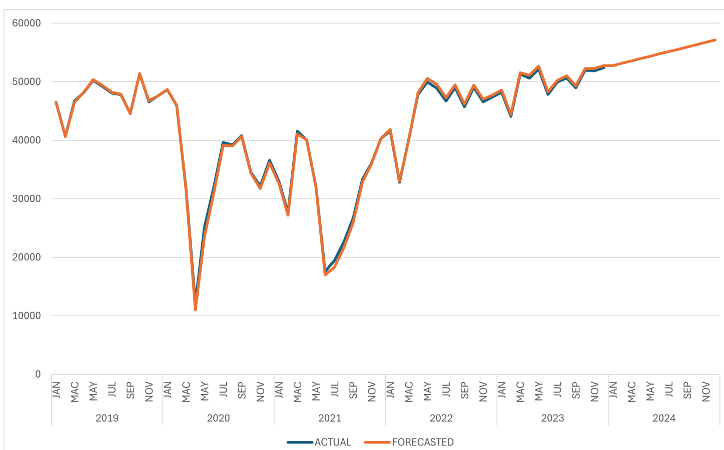


Figure 2: Graph Of Forecast Value Of Road Accident In Malaysia

Table 4: Forecasted Number of Accidents in 2024

Year	Month	Forecasted Number of Accidents in 2024, Y_t
2024	Jan	52,782.03367
	Feb	53,176.70636
	Mac	53,571.37905
	Apr	53,966.05174
	May	54,360.72443
	Jun	54,755.39712
	Jul	55,150.06981
	Aug	55,544.74250
	Sept	55,939.41519
	Oct	56,334.08788
	Nov	56,728.76058
	Dec	57,123.43327

4. Conclusion

The objective of this study was to model 60 data points from January 2019 to December 2023 of road accident data from Malaysia using Box-Jenkins ARIMA model and Univariate models, specifically SES and DES. The second objective invloved comparing error measurements, which is MAPE and RMSE, in order to choose the best model. Box-Jenkins was represented by ARIMA(1,1,1), and the best Univariate model was DES. According to the comparison, DES was the best model since it had the lowest RMSE and MAPE. As a result, from January 2024 to December 2024, DES was employed to forecast road accidents, with the eventual goal of reaching 60,000 cases each month. The study discovered that DES is the most accurate forecasting model, offering crucial information to policymakers and traffic experts to help them put plans into place that would lessen road accidents, which are a major cause of mortality in Malaysia. Besides, there are two recommendation machine learning algorithm which Naïve Bayes and Neural Network to come out with accurate result. The ability of neural networks to handle complex, non-linear interactions between variables like weather, traffic, and time of day makes them ideal for predicting traffic accidents (Gatarić et al., 2023).

References

Ali, M. (2023). 15 Most Common Types Of Road Accidents. https://www.hseblog.com/types-of-road-accidents/. [Accessed 22-03-2024].

Gatarić, D., Ruskic, N., Aleksić, B., Đurić, T., Pezo, L., Lončar, B., and Pezo, M. (2023). Predicting road traffic accidents-artificial neural network approach. *Algorithms*, 16:257.

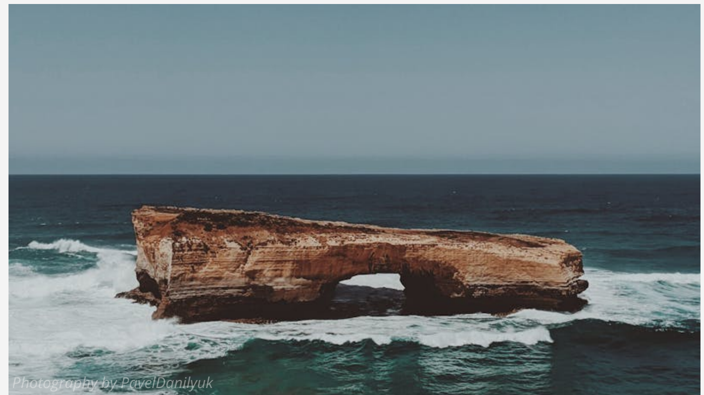
Gimino G., Rahim R., V. R. and T.A., Y. (2023). 915,874 road accidents recorded throughout 2021 and 2022, says Transport Ministry — thestar.com.my. [Accessed 22-03-2024].

Lazim, M. A. (2011). *Introductory business forecasting : a practical approach*. UiTM Press, Shah Alam, Selangor, 3rd ed edition. Includes 1 computer optical disc (4 3/4 in.).

Rosyid, H. A., Widiyaningtyas, T., and Hadinata, N. F. (2019). Implementation of the exponential smoothing method for forecasting food prices at provincial levels on java island. In *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pages 1–5.

World Health Organization, W. (2023). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. [Accessed 24-03-2024].

J O U R N A L O F
E X P L O R A T O R Y
M A T H E M A T I C A L
U N D E R G R A D U A T E R E S E A R C H



V O L U M E I I
O C T 2 0 2 4