PROCEEDINGS OF 5th MILTC 2023

MITRANS INTERNATIONAL LOGISTICS AND TRANSPORT CONFERENCE

20th DECEMBER 2023
DEWAN SIVIK, PETALING JAYA CITY COUNCIL

"TRANSPORT AND LOGISTICS IN EDUCATION FOR COMMUNITY"

Organised by

UNIVERSITI TEKNOLOGI MARA
Institut Pengangkutan Malaysia
MITRANS
Transporting Happiness Delivering Success

Co-Organisers

UNIVERSITI TEKNOLOGI MARA
ReNeU RESEARCH NEXUS UITM
ITL TRISAKTI
The Chartered Institute of Logistics and Transport
MBPJ MAJLIS BANDARAYA PETALING JAYA

PROCEEDINGS OF 2023 MITRANS INTERNATIONAL LOGISTICS AND TRANSPORT CONFERENCE (MILTC2023)

Organised by:

Malaysia Institute of Transport (MITRANS)

Co-Organised by:

Majlis Bandaraya Petaling Jaya (MBPJ)

Research Nexus UiTM (ReNeU)

Trisakti Institute of Transportation and Logistics (ITL)

The Chartered Institute of Logistics and Transport (CILT)

Editors:

S. Sarifah Radiah Shariff

Wan Mazlina Wan Mohamed

Nur Farizan Tarudin

Siti Aimi Sarah Zainal Abidin

Editors
S. Sarifah Radiah Shariff
Wan Mazlina Wan Mohamed
Nur Farizan Tarudin
Siti Aimi Sarah Zainal Abidin

Malaysia Institute of Transport (MITRANS)
Universiti Teknologi MARA, Shah Alam
40450, Shah Alam, Selangor
MALAYSIA.

Perpustakaan Negara Malaysia Cataloguing-in-Publication Data

Mitrans International Logistics and Transport Conference (5th : 2023 : Online)

PROCEEDING: 2023 MITRANS INTERNATIONAL LOGISTICS AND TRANSPORT CONFERENCE (MILTC2023): TRANSPORT AND LOGISTICS IN EDUCATION FOR COMMUNITY, 20 December 2023, Virtual Conference, Dewan Sivik, Petaling Jaya City Council, Selangor / Organised By : Malaysia Institute of Transport (MITRANS) Majlis Bandaraya Petaling Jaya (MBPJ) Trisakti Institute of Transportation and Logistics (ITL) Research Nexus UiTM (ReNeU) The Chartered Institute of Logistics and Transport (CILT); Editor: S. Sarifah Radiah Shariff, Wan Mazlina Wan Mohamed, Nur Farizan Tarudin, Siti Aimi Sarah Zainal Abidin
Mode of access: Internet

e ISBN : XXX-XXX-XXXXX-X-X

1. Business logistics--Congresses. 2. Logistics--Congresses. 3. Transportation--Congresses. 4. Government publications--Malaysia. 5. Electronic books. I. S. Sarifah Radiah Shariff. II. Wan Mazlina Wan Mohamed. III. Nur Farizan Tarudin. IV. Siti Aimi Sarah Zainal Abidin. V. Institut Pengangkutan Malaysia. VI. Majlis Bandaraya Petaling Jaya. VII. Neksus Penyelidikan, UiTM Shah Alam. VIII. Institut Transportasi dan Logistik Trisakti. IX. The Chartered Institute of Logistics and Transport. X. Title. 658.5

PREFACE

The MITRANS International Logistics & Transport Conference (MILTC) has been an annual event since its inception in 2016. The fifth edition, MILTC2023, took place on 20th December 2023, at the Dewan Sivik, Petaling Jaya City Council, Selangor. The conference was organized in collaboration with several partners, including Majlis Bandaraya Petaling Jaya (MBPJ), Research Nexus in Transport and Logistics, ReNeu UiTM; Trisakti Institute of Transport and Logistics (ITL) from Indonesia; and the Chartered Institute of Transport Malaysia (CILTM) Selangor Section.

MILTC2023 featured two distinguished keynote speakers from academia and industry who shared their knowledge and experiences on the first day of the event. These speakers delivered insightful presentations that enriched the conference.

The conference received a total of 43 papers, which were selected for oral presentations during the two- day event. Authors from both industries and academia from Malaysia and Indonesia contributed to the conference by presenting their research findings, innovations, and technological advancements in the field of Transportation, Logistics, and Supply Chain.

The abstracts of all papers, along with a selection of full papers, are included in the conference proceedings. This publication aims to disseminate knowledge and contribute to the advancement of the field of Transport, Logistics, and Supply Chain within education while serving the community.

The organizing committee would like to express their sincere gratitude to the keynote speakers, secretariats from both MITRANS and MBPJ Learning City team (including academics, support staff, and students), the reviewers, and especially the authors and participants who made MILTC2023 a memorable event. The committee also acknowledges the support of the top management of MITRANS and UiTM in making the conference possible.

It is the committee's sincere hope that the Proceedings book will serve as a valuable resource for the advancement of knowledge in the field of Transport, Logistics, and Supply Chain.

Chief Editor
Associate Professor Ts. Dr. S Sarifah Radiah Shariff
Deputy Director
Malaysia Institute of Transport (MITRANS)
Universiti Teknologi MARA (UiTM)

TABLE OF CONTENT

MESSAGE FROM THE ADVISOR OF MILTC 2023

Assalamu'alaikum warahmatullahi wabarakatuh, Peace Be Upon You.

Dear Distinguished Guests and Conference Participants,

We sincerely congratulate the success of the 5th MITRANS' International Logistics and Transport Conference 2023 (MILTC2023) to the Malaysia Institute of Transport (MITRANS), Universiti Teknologi MARA (UiTM), and conference committee members. Emerging technological advances are having a huge influence on supply chain management, logistics, and even transportation sustainability. The logistics industry benefits the most from its broad usage of manual methods and vast amount of data to retain. As a result, adding new technologies such as artificial and augmented intelligence, advanced analytics, and automation appears to be fashionable, original, and intriguing.

In keeping with our conference subject," Transport and Logistics in Education for Community", we are aware that technology is advancing at a rapid pace, resulting in new solutions and innovations. Industries that specialize in these sectors would have to adapt to the new expectations and standards or risk falling behind. bear in mind, however, that new business models and industry participants have also had a key role in contributing to the challenges, so bear in mind that it is not only new technology that will influence these sectors in the future.

I would like to express my heartfelt appreciation to all participation and partnerships in making this event a success. Not to mention our four distinguished keynote speakers: Yang Berbahagia Prof Ir. Dr. Ahmad Kamil Arshad, Director of Institut Kejuruteraan Infrastruktur dan Pengurusan Mampan (IIESM), UiTM Shah Alam and Tuan Tpr. Lee Lih Shyan (Pengarah Perancangan Pembangunan of Petaling Jaya Council (MBPJ)).

Furthermore, we hope that this platform helped to the development of a strong synergy between academics and business. Last but not least, the conference would not be possible without the papers submitted by authors, thus we appreciate their efforts and participation in the MILTC2023. Finally, on behalf of the organizing committee, I really hope that this one dayl conference will be fruitful and engaging for all of you, motivating you to continue doing high-quality research.

Thank you

Professor Ts. Dr. Norazah Abd Rahman
Deputy Vice-Chancellor (Research & Innovation)
Universiti Teknologi MARA (UiTM)

MESSAGE FROM THE CHAIRPERSON OF MILTC 2023

Assalamualaikum Warahmatullahi Wabarakatuh and Good Day.

As the conference chairperson and on behalf of Malaysia Institute of Transport (MITRANS), Universiti Teknologi Mara (UiTM), I would like to extend a warm appreciation to all speakers, co-organizers, and participants of the 5th MITRANS International Logistics and Transport Conference.

We would like to offer our heartfelt appreciation to all of the attendees who have responded positively to this conference. For this conference, we got 43 papers (from both international and local participants, from academics as well as industry).

MILTC 2023 is an MITRANS project to encourage research findings sharing among players in the transportation, logistics, and halal supply chain through oral presentations and publications. The goals are to foster research in these disciplines and to allow the sharing of new ideas. It comprises plenary sessions, keynote speeches, and oral presentations on various themes.

We think that everyone attending this conference shares the goal of researching and comprehending current and cutting-edge topics in order to remain at the forefront of knowledge. As a result, we really hope that the knowledge gathered from this conference will serve as a springboard for further investigation into the possibility of new research agendas for the benefit of humanity. Thus, it delivers high-quality scientific information as well as essential networking opportunities for scholars from across the world.

Finally, we'd want to take this occasion to thank all MILTC 2023 committee members for their time and effort. MILTC 2023 would not be a reality without their endless effort and hard work. Many thanks also to the co- organizer, UiTM management, and everyone else who helped make this conference a success.

Finally, we wish you a successful career ahead. Thank you.

Associate Professor Madya Ts. Dr. S Sarifah Radiah Shariff
Chairperson Of MILTC 2023,
Malaysia Institute of Transport (MITRANS) Universiti Teknologi MARA (UiTM)

ORGANISING COMMITTEE

Advisor

Professor Ts. Dr. Norazah Abd Rahman

Chairperson

Associate Professor Ts. Dr. S Sarifah Radiah Shariff

Deputy Chairperson

Associate Professor Ts. Dr. Wan Mazlina Wan Mohamed

Secretary

Ts. Nur Farizan Tarudin

Secretary II

Pn. Fatimah Ayub

Treasurer

Pn. Zanirah Zamrud

Secretariat Member (Promotion & Website)

Ts. Dr. Abdul Khabir Rahmat

Pn. Dayang Siti Nurizeyan Mohd Imbran

Secretariat Members (Registration and Usherer)

Pn. Fatimah Ayub

Pn. Surniwati Nasrun

Pn. Zanirah Zamrud

Secretariat Member (Award)

Dr. Siti Ayu Jalil

Ts. Dr. Zaimy Johana Johan

Secretariat Members (Technical, Logistics & Multimedia)

En. Suib A Wahab

En. Langga Anak Rimon

En. Faizal Mohd Amin Sharifuldin

Secretariat Member (Proceeding)

Dr. Siti Aimi Sarah Zainal Abidin

Secretariat Members (Gift and Certificate)

Pn. Fatimah Ayub

Pn. Zanirah Zamrud

Secretariat Members from MBPJ Learning City

Ts. Hjh Jafizatul Azlinda Binti Jaafar

Pn. Nor Shalizawati Binti Hj. Baharudin

En. Zainal Abidin Bin Rahim

En. Bonny Anak Eddy

Pn. Nurul Ain Binti Hasan

Pn. Sahlawati Binti Abdul Ghany

En. Imran Shawal Bin Samsudin

Pn. Nurul Arissa Binti Mohd Sabariman

Puan Noraszlina Mohd

Secretariat Members (Photographer)

En. Mohd Helmi Al Bakri Mohd Hairi

En. Wan Azlan Mat Isa

Secretariat Members (Presentation & Others)

En. Akmal Haziq Ahmad Aizam

Pn. Dayang Siti Nurizeyan Mohd Imbran

Pn. Siti Nur Akhluma Jaafar

En. Syed Shaufiq Shahredza Syed Baharuddin

En. Amar Amjad Abdul Wahid

Pn. Elanis Rachel Raymond

Pn. Fatin Najwa Mohd Nusa

Pn. Juliana Hamka Kamaroddin

Pn. Nurul Hazwani Hamzah

Pn. Zanirah Zamrud

Emcee (MC)

En. Saiful Bahari Bin Ab Azais

En. Nursyazwan Muhamad Isa

FULL PAPERS

* Please click on the title to access each paper directly.

PAPER 10

USING PREDICTIVE ANALYTICS TO SOLVE A NEWSVENDOR PROBLEM

S. Sarifah Radiah Shariff[1], Hady Bin Hud[2*]

[1]*Malaysia Institute of Transport (MITRANS), Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia*
[2]*MISI Universiti (formerly known as Malaysia Institute for Supply Chain Innovation), Persiaran Tebar Layar, Bukit Jelutong, 40150 Shah Alam, Selangor, Malaysia*

[*]*Corresponding author: hhud@misi.edu.my*

Abstract

Purpose
Forecasting accuracy is an ongoing challenge for many companies. The pandemic has resulted in unpredictable customers' purchasing patterns post-pandemic has rendered heuristic-based forecasting large forecast errors, which leads to poor decision-making. It is debatable whether such behaviour may persist long after the pandemic ends or settle down at a new normal level. The objective of this research is to forecast the daily demand of a perishable product during the new norm post-pandemic, ensuring minimal unsold items are discarded as waste.

Findings
This research makes two key observations. Firstly, an algorithm with good performance metrics over a small data set collection may obtain worse results when the data set collection is extended. The best algorithm will not be the same for all the data sets. Secondly, in solving every Machine Learning problem, there is no one algorithm superior to other algorithms. Every algorithm makes its own respective prior assumptions about the relationships between the features and target variables, which create different types and levels of bias. The assumptions adopted in the Decision Tree Model and K-Nearest Neighbour are derived from symbolic artificial intelligence and data mining, whilst the assumptions in the Artificial Neural Network are derived from the connectionist approach.

Practical implications
In terms of managerial implications, the findings in this research help to frame the adoption of a more advanced analytical approach to forecasting, using a Machine Learning algorithm, in solving a newsvendor problem. The unpredictability of customers' purchasing patterns post-pandemic has rendered heuristic-based forecasting large forecast errors. This research attempts to solve the problem statement on how to forecast the daily optimal quantity of a perishable product during the new norm post-pandemic, ensuring minimal unsold items are discarded as waste.

Originality/value
Overall, this research provides initial insights into adopting Machine Learning algorithms in making better-informed managerial decisions among SMEs in Malaysia.

Keywords: Machine Learning Algorithms, Demand Forecasting, Perishable Items, Newsvendor Problem, Inventory Management.

Introduction

SME plays important role in Malaysia's economic transformation process, fostering growth, employment, and income. This has resulted in significant economic and social progress for several decades, thus facilitating a transition from a low-income to a middle-income nation. The role of SMEs will become increasingly critical, not only as enablers of growth by providing support to large firms but also as a driver of economic growth. SME development is also important in achieving a more balance and inclusive growth, by addressing the bottom 40% of the income pyramid, which include microenterprises (SME Corporation Malaysia, 2012).

Table 1 outlined how SMEs are classified in Malaysia. Two criteria are adopted (i.e., the annual total sales turnover and the number of full-time employees), with further juxtaposed into two dimensions (i.e., the manufacturing industry, and the services and others).

1. SME is classified as Micro when either the annual sales turnover is less than RM0.3 million or has less than five employees. This classification is applicable to both SMEs in the manufacturing industry as well as in the services and others.
2. SME is classified as Small when, in the case of manufacturing industry, either the annual sales turnover is from RM0.3 million to less than RM15 million or has employees from 5 to less than 75 people. In the case of services and others, either the annual sales turnover is from RM0.3 million to less than RM3 million or has employees from 5 to less than 30 people. The threshold for the manufacturing company is higher than the threshold for the services and others.
3. SME is classified as Medium when, in the case of manufacturing industry, either the annual sales turnover is from RM15 million to not exceeding RM50 million or has employees between 75 and not exceeding 200 people. In the case of services and others, either the annual sales turnover is from RM3 million to not exceeding RM20 million or has employees between 30 and not exceeding 75 people. The threshold for the manufacturing company is higher than the threshold for the services and others.

Table 1: SME Classification in Malaysia ((SME Bank, 2022)

Category	Manufacturing	Services and Others
Micro	Sales turnover of less than RM300,000 OR Employee: Less than 5	Sales turnover of less than RM300,000 OR Employees: Less than 5

Category	Manufacturing	Services and Others
Small	Sales turnover from RM300,000 to less than RM15 Million OR Employees: From 5 to less than 75	Sales turnover from RM300,000 to less than RM3 Million OR Employees: From 5 to less than 30
Medium	Sales turnover from RM15 Million to not exceeding RM50 Million OR Employees: From 75 to not exceeding 200	Sales turnover from RM3 Million to not exceeding RM20 Million OR Employees: From 30 to not exceeding 75

In aggregate, SME is a major contributor to the economy of Malaysia. Figure 1 indicates SMEs is contributing 38.2% of the GDP of Malaysia in 2020 (declined marginally when compared to 2019 of 38.9%). In terms of absolute value-added amount, SMEs generated RM512.8 billion in 2020 (marginally declined when compared to 2019 of RM553.5 billion). Historically, the annual growth of SMEs' contribution to GDP is usually higher than the annual GDP growth of Malaysia. In 2020, this trend does not hold true anymore, where the annual growth of SMEs' contribution to GDP has registered at negative 7.3% (while the annual GDP growth of Malaysia is at negative 5.6%). The last time a similar pattern took place was in 2003.

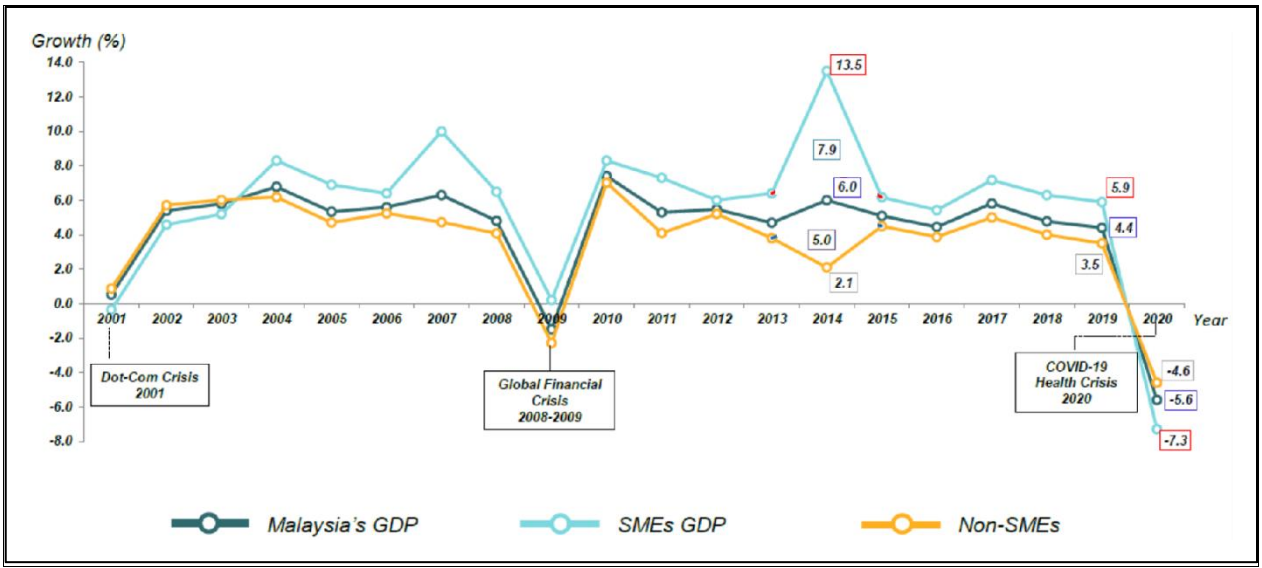


Figure 1: Annual GDP Growth at Constant Prices (Source: Department of Statistics, 2021)

Many SMEs have been grossly underestimated the potential benefits of SCMP, while others viewed SCMP as additional operating expenses (Chin, Hamid, Rasli, & Baharun, 2012). Other scholars highlighted that the majority of SMEs in Malaysia have insufficient knowledge of SCMP (Rahman, Wasilan, Deros, & Ghani, 2011). It is observed that many researchers often time fails to integrate the perspective of individual practitioners in developing research design (Rynes, Bartunek, & Daft, 2001). Therefore, this study is motivated to address the SMEs' research-practice gap, focusing on the fifth component of the SCMP. All demand planners, within the supply chain industry, face a challenging task in formulating demand forecasting. This is particularly more acute within SMEs. The shift in customer behavior during the new norm post-COVID-19 resulted in inconsistent demand patterns, making it harder to generate baseline forecasts based on past consumption trends. Perishable products have expiration dates, turning bad if not consumed within a certain period. After the expiration dates, it has no value and discarded as waste. Hence, a bread producer will produce enough supply for one period, limiting the amount of inventory that can be discarded as waste. Disappointed loyal customers would likely be migrated to competitors, leading to losing market share and damaging corporate reputation. In this study, the optimal quantity of perishables (i.e., "Roti Burger") ensures minimal unsold items discarded as waste.

Methodology

The data is based on the actual production data between 1 January 2022 and 31 July 2022. There was no production data during the Movement Control Order Period between March 2020 and February 2021. The sponsor company was closed for operations to conserve cash in minimizing the economic knockout due to the lockdown restriction. As the state of Johor transitioned to Phase 4 of the National Recovery Plan on 25 October 2022, the restriction on social gathering has been removed and all economic sectors will be reopened (Azil, 2021). The sponsor company re-commenced its production operations in November 2021, with demand for "Roti Burger" intermittently picked-up in November and December 2021. As this demand trend is not a representative, it was excluded for the purpose of this study, and to focus production data from January 2022 onwards.

Upon completion of data collection, steps are taken to enhance data quality ready for onward processing. Data is split into testing data and training data, so as to reduce the degree of data dispersion (Kisvari, Lin, & Liu, 2021).

1. Formatting the data to ensure all variables are within the same attribute (e.g., all dates and production volumes are written in the same format).
2. Cleaning the data to ensure any missing data is replaced with mean values as well as removing any duplicates and outliers, especially if upon consultation with domain expertise indicates the immateriality of the duplicates and outliers for the model.
3. Encoding textual data to numbers (i.e., assign integers to replace Sunday, Monday, etc.)) as models are not able to process textual data.
4. Sampling the data into a smaller representative sample for faster modeling compared to the whole dataset. The algorithm in this study specifies train data size of 75%, which means that the algorithm learns to build a model based on 75% of the data for the period of January to July 2022, while the remaining 25% will be the test data used to evaluate the generalization performance of the model.
5. Scaling the pre-processed data into a more manageable scale suitable for a specific machine learning algorithm. This study aims to analyze the relevant machine learning algorithms to forecast daily perishable "Roti Burger" and measure the performance

metrics using testing data. The algorithm with the most favorable performance metrics will be selected for the deployment phase where predictions will be based on the actual production data.

Using Python Programming Language with Panda library (to perform data cleaning and exploratory analysis), this study generates Descriptive Statistics, a statistical method used to describe the following three basic features of the dataset:

1. The central tendency measurement, describing the center of the dataset (mean, median and mode).
2. The variability (or spread) measurement, describing the dispersion of the dataset (variance and standard deviation).
3. The frequency distribution measurement, describing the occurrence of data within the dataset (count).

Results of the multiple machine learning techniques are compared and assessed against the two pre-identified performance metrics (i.e., MAE and RMSE). This study identified the Decision Tree Model as the most suitable forecasting algorithm as it has the lowest forecast error in both performance metrics.

Results and Discussion

Table 2 refers to the Descriptive Statistics of the actual production data for the period between 1 January 2022 and 31 July 2022. There are 181 observations in this study. Although a sample size of 30 observations is the minimum size for the Central Limit Theorem and are adequate to represent the data distribution, some practitioners recommend a minimum of 50 observations for a more representative distribution with a histogram (Chang, Huang, & Wu, 2006). As the higher the sample size, the more likely the sample mean, and sample standard deviation will be more representative to the population mean and population standard deviation. Hence, 181 observations can be considered a sufficiently large dataset. Analyzing the central tendency and the frequency distribution measurements of the dataset indicates the following observations:

1. On average daily basis, there were less “Roti Burger” Sold to customers than were produced (an average of 4,606 pieces per day were sold as compared to an average of 4,613 pieces per day were produced). This has resulted, on average, 33 pieces per day were discarded.
2. Since the median of “Roti Burger” Sold to customers is 4,980 pieces is not identical to the mean of 4,607 pieces, indicates that it is not in a symmetric distribution (i.e., not a normal distribution).

Alternatively, the dataset can be dissected into day-to-day basis as per Figure 2. It is evidenced from the Boxplot chart above that the boxplot for Saturday is the thinnest (indicating that Saturday data is the least dispersed due to seasonality factors). As the days progressed, the data became more dispersed and the boxplots gradually become thicker, with Thursday is the thickest. The presence of outliers differs from day to day during 2022. There is no data for Friday as the sponsor company is closed for operations. The unpredictable customer demand post-pandemic makes it difficult to forecast the daily production of “Roti Burger”. Since a heuristic approach has been adopted to forecast daily production, there are evidence of overproduction, resulted in higher wastage. It is worth investigating the non-seasonality fluctuations in order to minimize discarded “Roti Burger”.

Table 2: Descriptive Statistics

	Day	Month	RotiBurgerProduced	RotiBurgerSold	RotiBurgerDiscard
count	181.0	181.0	181.0	181.0	181.0
mean	3.7	4.0	4,613.3	4,606.6	33.2
std	2.0	2.0	1,339.4	1,350.1	32.6
min	1.0	1.0	-	-	-
25%	2.0	2.0	5,000.0	4,860.0	-
50%	4.0	4.0	5,000.0	4,980.0	30.0
75%	5.0	6.0	5,000.0	5,080.0	60.0
max	7.0	7.0	5,000.0	5,460.0	200.0

Alternatively, the dataset can be dissected into day-to-day basis as per Figure 2. It is evidenced from the Boxplot chart above that the boxplot for Saturday is the thinnest (indicating that Saturday data is the least dispersed due to seasonality factors). As the days progressed, the data became more dispersed and the boxplots gradually become thicker, with Thursday is the thickest. The presence of outliers differs from day to day during 2022. There is no data for Friday as the sponsor company is closed for operations. The unpredictable customer demand post-pandemic makes it difficult to forecast the daily production of “Roti Burger”. Since a heuristic approach has been adopted to forecast daily production, there are evidence of overproduction, resulted in higher wastage. It is worth investigating the non-seasonality fluctuations in order to minimize discarded “Roti Burger”.

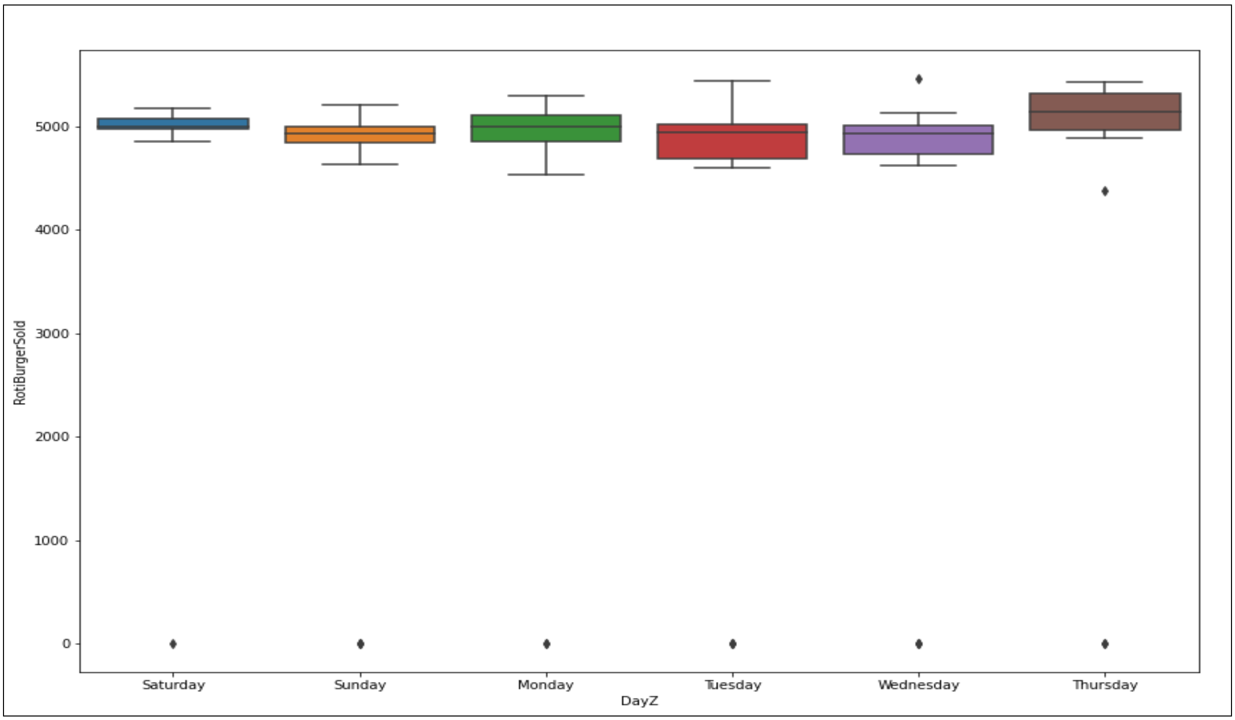
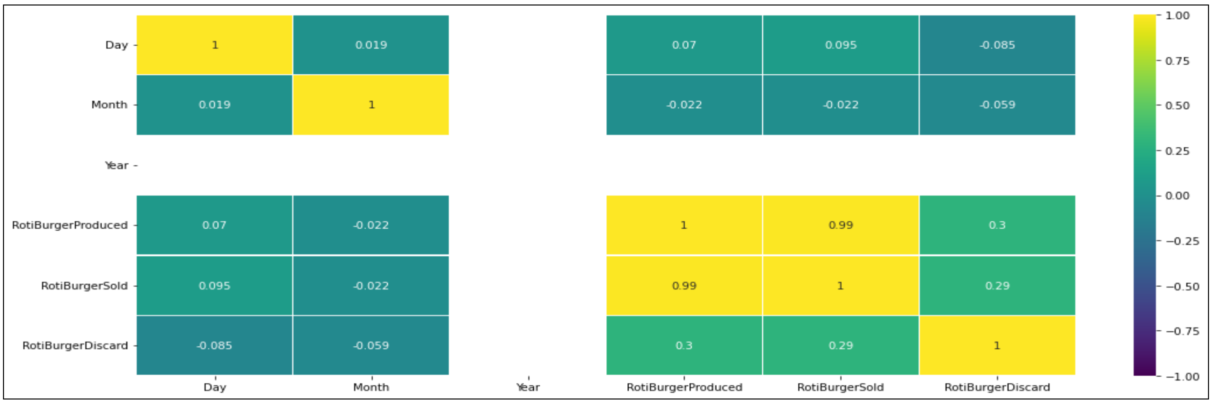


Figure 2: Boxplot on Daily Pieces Sold (Day-on-Day Basis)

This study also performs analysis on the Pearson Correlation Coefficients to determine whether there is linear relationship between variables in the dataset (refer Table 3). Pearson Correlation Coefficients measure the strength and direction of linear relationship between two variables, based on a scale between -1 and +1. A perfect negative correlation has a correlation value of -1 (i.e., one variable increases as the other variable decrease). A perfect positive correlation has a correlation value of +1 (i.e., one variable increases as the other variable increase). A complete absence of correlation is represented by 0. The underlying assumption when computing the Pearson Correlation Coefficient is that the variables are independent to each other. Notwithstanding of the above, a Pearson Correlation Coefficient does not imply causation (Conn, 2017). There are several factors influencing the daily "Roti Burger" production. Table 3 provided a summary of all Pearson Correlation Coefficients. There is weak correlation between "Roti Burger" Sold and Day as well as between "Roti Burger" Sold and Month (all of which have correlation value of almost 0 respectively). In this sense, a weak correlation indicates a weak linear relationship between these variables. This further validates the non-linear relationship between "Roti Burger" Sold and Day as well as between "Roti Burger" Sold and Month that have been established in earlier sections of this study. Although there is almost perfect positive correlation between "Roti Burger" Produced and "Roti Burger" Sold (a positive correlation value of 0.99), both variables are not independent to each other, violating the underlying assumption in the correlation computation. Likewise, the moderate positive correlation between "Roti Burger" Sold and "Roti Burger" Discards (with a positive correlation value of 0.29) violates the correlation computation assumption as both variables are not independent to each other. In summary, "Roti Burger" sold is influenced by when it was sold (either day or month), with the remaining variables have been ruled out due to assumption violation (i.e., “Roti Burger” Production and “Roti Burger” Discards).

Table 3: Correlation Table



The Machine Learning Pipeline approach requires all data to be pre-processed prior to the splitting of train data and test data. All categorical features (i.e., Date, Day, and Month) are converted into numerical features. There were no missing data in the dataset. The split between train data and test data is performed based on 75%: 25% ratio. The purpose of the split is to ensure that the algorithm is not trained on all the available data and test how the data performs on unseen data. If all the data is used as train data, then the outcome will be an overfitted algorithm (which perform poorly on unseen data). Overfitting is a term used in statistics where the machine learning model fits the training data too well and as a result, the model fails to fit on additional new data, and this may affect the accuracy of predicting future observations. Algorithms with high variance in the train data are often too complex and lead to overfitting, whilst algorithms with high bias are often too simple and lead to underfitting. The bias of an algorithm refers to the error that comes from the potentially wrong prior assumptions in the algorithm. These assumptions cause the algorithm to miss important information about the relationship between the features and target variables. This study is of the view that the 75%:25% split is deemed optimal to have enough bias to avoid simply memorizing the train data and enough variance to actually fit the patterns in the train data. With 181 observations in the dataset, train data and test data contain 135 and 46 observations respectively. Machine learning algorithms perform better when numerical input variables are scales to a standard range. In this study, MinMaxScaler is used for scaling the pre-processed data into the range [0,1], being the default range for normalization. The MinMaxScaler preserves the shape of the original distribution, and it doesn’t meaningfully alter the information embedded in the original dataset. Key findings from data analysis using Decision Tree Model, KNN, and Neural Network are in the ensuing sections.

Based on Table 4, the Decision Tree Model has lower scores of both performance metrics (MAE and RMSE) as compared to KNN. The most suitable time series forecasting algorithm is the Decision Tree Model as it has the lowest error (as measured by both MAE and RMSE). One study has observed that some machine learning techniques with a good average performance over a small data set collection could obtained significantly worse results when the data collection is extended. Conversely, machine learning techniques with suboptimal performance on the small data collection will be not so bad when more data sets are added-in (Macia & Bernadó-Mansilla, 2014). This is in total alignment with the No Free Lunch Theorem where the best machine learning algorithm will not be the same for all the data sets (Fernández-Delgado, 2014). In the context of machine learning, this scenario fits the description of the black swan paradox. Every machine learning algorithm makes its own respective prior assumptions about the relationship between the features and target variables which create different types and levels of bias. The assumptions adopted in Decision Tree Model and KNN are derived from symbolic artificial intelligence and data mining, whilst the assumption in Neural Networks derived from the connectionist approach. The performance metrics of Decision Tree Model is better than KNN and Neural Network for the data set in this study, but every algorithm has advantages and disadvantages when aligning the prior assumptions with the data set.

Table 4: Summary of Performance Metrics

Type of Indicators	Decision Tree Model	KNN	Neural Network
MAE	148.48	301.93	4781
RMSE	213.30	765.73	4893

Table 5 compares the actual “Roti Burger” sold during the month of August 2022 against the forecasted figures, where on aggregated basis, there were 129,980 pieces were actually sold as compared to 130,010 pieces that were forecasted to be sold as generated using the Decision Tree Model. In terms of performance metrics, MAE for the August 2022 is 179.63 while RMSE is 225.73, which are consistent trend with the MAE and RMSE in Table 4 (of 148.48 and 213.30 respectively). In the event where the Decision Tree Model is deployed, constant monitoring and testing need to be done on the production data, particularly on the performance metrics of MAE and RMSE. This is necessary so as to track possible degradation of the algorithm model.

Table 5: Actual August 2022 Data Against Forecast

Date	Actual Roti Burger Sold a	Forecast Roti Burger Sold f	Error a-f	| Error | | a-f |	Error $(a-f)^2$
1-Aug-22	4,600	4,770	-170	170	28,900
2-Aug-22	5,370	4,980	390	390	152,100
3-Aug-22	4,980	5,100	-120	120	14,400
4-Aug-22	5,020	5,010	10	10	100
6-Aug-22	4,980	4,790	190	190	36,100
7-Aug-22	4,980	5,110	-130	130	16,900
8-Aug-22	4,930	5,030	-100	100	10,000
9-Aug-22	5,140	5,220	-80	80	6,400
10-Aug-22	4,720	4,820	-100	100	10,000
11-Aug-22	5,130	5,430	-300	300	90,000
13-Aug-22	5,060	4,850	210	210	44,100
14-Aug-22	5,090	5,000	90	90	8,100
15-Aug-22	4,700	4,980	-280	280	78,400
16-Aug-22	5,210	5,000	210	210	44,100
17-Aug-22	5,090	5,000	90	90	8,100
18-Aug-22	5,000	5,460	-460	460	211,600
20-Aug-22	4,990	4,970	20	20	400
21-Aug-22	5,000	4,920	80	80	6,400
22-Aug-22	4,980	5,070	-90	90	8,100
23-Aug-22	4,930	4,990	-60	60	3,600
24-Aug-22	5,080	4,600	480	480	230,400
25-Aug-22	4,980	4,930	50	50	2,500
27-Aug-22	4,870	4,640	230	230	52,900
28-Aug-22	5,140	5,290	-150	150	22,500
29-Aug-22	4,680	5,080	-400	400	160,000
30-Aug-22	5,330	4,970	360	360	129,600
31-Aug-22	0	0	0	0	0
Total	129,980	130,010	-30	4,850	1,375,700

Conclusion

The findings in this study were valuable to the sponsor company because it gives the ability to make better-informed managerial decisions, particularly by moving away from using heuristics (i.e., mental shortcuts) as a basis for time series forecasting towards a more systematic basis approach. Whilst using any rule-of thumb approach can figure out a solution to the daily production problem faster, it can also lead to inaccurate judgement that leads to suboptimal output with inconsistent forecast accuracy. Due to its perishable nature, "Roti Burger" can only be sold within few days from the production date. Any unsold item after this period is considered obsolete, discarded, expensive and making reordering impractical. On the other hand, running out-of-stock leads to an opportunity loss of revenue as customers are not able to buy the bread they wish for.

Furthermore, the shift in customer behaviour during the new norm post-COVID-19 resulted in inconsistent demand patterns, making it harder to generate baseline forecasts based on past consumption trends. In many circumstances, demand forecasting is performed using linear regression, but the shortcoming of this approach is that data analyzed often exhibit some non-linearity that cannot be captured by a linear regression algorithm. The Machine Learning Pipeline approach offers a flexible unified approach, with better understanding of the key drivers of daily "Roti Burger" sold, particularly where nonlinear relationship exists. It allows calibration to increase our understanding of the landscape characteristics that shape the daily "Roti Burger" sold. Combined with dynamic pricing and promotions, managers will be able to identify price sensitivity among customer segments where the managers can stimulate incremental consumer purchases, yielding future sales growth.

References

Azil, F. (2021, October 22). *Johor, Terengganu masuk Fasa 4 PPN bermula 25 Oktober*. Retrieved from Astro Awani: https://www.astroawani.com/berita-malaysia/johor-terengganu-masuk-fasa-4-ppn-bermula-25-oktober-perdana-menteri-326795

Chang, H., Huang, K., & Wu, C. (2006). Determination of sample size in using central limit theorem for weibull distribution. *International Journal of Information and Management Sciences, Vol. 17, No. 3*, 153-174.

Chin, T., Hamid, A., Rasli, A., & Baharun, R. (2012). Adoption of supply chain management in SMEs. *Procedia-Social and Behavioral Sciences*, 614-615.

Fernández-Delgado, M. C. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research, 15(1)*, 3133-3181.

Kisvari, A., Lin, Z., & Liu, X. (2021). Wind power forecasting–A data-driven method along with gated recurrent neural network. *Renewable Energy, 163*, 1895-1909.

Macia, N., & Bernadó-Mansilla, E. (2014). Towards UCI+: a mindful repository design. *Information Sciences, 261*, 237-262.

Rahman, M., Wasilan, H., Deros, B., & Ghani, J. (2011). Barriers of SCM in SMEs. *Applied Mechanics and Materials*, 44 - 47.

Rynes, S. L., Bartunek, J. M., & Daft, R. L. (2001). Across the great divide: Knowledge creation and transfer between practitioners and academics. *Academy of management Journal, 44(2)*, 340-355.

SME Bank. (2022). *SME Definition*. Retrieved from SME Bank: https://www.smebank.com.my/en/

SME Corporation Malaysia. (2012). *SME Masterplan 2012-2020 : Catalysing Growth and Income.* Kuala Lumpur: SME Corporation Malaysia.