Information Management and Business Review (ISSN 2220-3796)
Vol. 16, No. 3, pp. 201-209, Sep 2024

Predicting Consumer Behavior in E-Commerce Using Decision Tree: A Case Study in Malaysia

Nurul Ain Mustakim[1]*, Maslina Abdul Aziz[2], Shuzlina Abdul Rahman[2]
[1]Faculty of Business & Management, Universiti Teknologi MARA, Cawangan Melaka Kampus Bandaraya Melaka, Melaka, Malaysia
[2]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Shah Alam, Malaysia
ainmustakim@uitm.edu.my*, maslina_aziz@uitm.edu.my, shuzlina@uitm.edu.my
Corresponding Author: Nurul Ain Mustakim

Abstract: Understanding and predicting consumer behavior will help e-commerce businesses improve customer satisfaction and devise better marketing strategies. This study is intended to explore the use of decision tree algorithms in predictions of consumer purchase behavior in the e-commerce platform in Malaysia. Comparing the performances of J48, Random Tree, and REPTree decision tree models using an online shopper dataset collected by surveying 560 Malaysians, on various aspects like accuracy, precision, recall, and F1 score. Results indicate that the highest accuracy has been achieved with the Random Tree algorithm, outperforming J48 and REPTree. The results will, therefore, form the basis upon which e-commerce can restrategize its marketing programs for better customer engagement. This is an important study in that it shows the efficacy of applying a decision tree algorithm to understand customer behavior in the context of Malaysia and adds to the growing body of knowledge in predictive analytics in e-commerce.

Keywords: *Consumer Behaviour, E-Commerce, Decision Trees, Predictive Analytics, Malaysia*

1. Introduction

E-commerce in Malaysia is growing, and this generally means increased digitization and a shift in consumer behavior. To be able to compete in this fast-growing market, companies should be sure to rely on machine learning and AI. The technology will help explain and predict consumer preferences and behavior while shopping. According to Digital Malaysia 2020, half of Malaysians shop online, and 82.9 percent of mobile users use their phones to shop online. It is expected, in turn, that from 2023 to 2027, an additional 2.9 million people will join Malaysia's e-commerce market. Because of the unique cultural diversity represented by the Malay, Chinese, and Indian communities, Malaysia comprises a complex and enriched environment in which research into consumer behavior in e-commerce can be conducted. This diversity influences purchasing patterns, preferences, and motivations that need understanding in the Malaysian context. The takeaway from such studies can be used to inform focused marketing strategies that allow businesses to align operations with what is expected by consumers. This may also assist in helping policymakers develop programs that continue to support growth in e-commerce, particularly in underserved areas.

Even though Malaysia is considered an advanced country in the domains of e-commerce and machine learning, there exist some unique challenges that it faces in adapting to these technologies. Regarding Loh and Teoh (2021) and Vachkova et al. (2023), machine learning in Malaysia is still in its very infant stages. The existing technological facilities need to be used very judiciously so that economic growth can take place along with international expansions, but in this regard also, there exists a lot of disparity. Still, formidable challenges involve heavy investment in building up a strong technology infrastructure, training the workforce, and privacy and security management of users' data. Castillo and Taherdoost (2023) continue to explain that although big global companies such as Amazon, Alibaba, and Rakuten remain on top using AI in big data and chatbots, Malaysian firms lack the latest technology, restrictive regulations, and human resource deficiencies in developing required digital skills. These questions indicate the divide between global developments and local adaptation, hence caveating the fact that an approach tailor-made to fully uncover AI and ML integration into the Malaysian e-commerce market needs to be put into the sling of affairs.

One of which is that many studies (e.g., Kiew et al., 2021; Kuswanto et al., 2020; Alhakim et al., 2023; Pe et al., 2020; Sarangapani et al., 2023) have found several factors that influence consumer behavior online, including attitude, motivation, perceived risk, and demographic variables. Although these studies have been able to shed much valuable insight, they have often relied upon descriptive models of consumer behavior that capture a snapshot at any given time and fail to examine how these factors dynamically interact over time to predict future behavior. Most research conducted in Malaysia has not taken full advantage of machine learning approaches yet for more accurate results. However, much has been looked upon in terms of the merits and demerits of predicting consumer behavior, a thing that has created a gap in research work targeting Malaysian consumers. This study will specifically set a goal of understanding local consumer behavior in the Malaysian context as a way of addressing this gap and reacting to the fast-rising e-commerce market nature. Previous studies only considered general online consumer behavior without fully considering the unique features of the Malaysian market. This paper applies the machine learning approach in decision tree analysis, targeting the study of analyzing and predicting Malaysian consumer behavior. In turn, decision trees are handled as very effective means to catch other complex non-linear data patterns in large data series and provide an organized and interpretable model for decision-making. With specific reference to the Malaysian online market, uniqueness springs from the cultural diversity in which it operates, the rapid technology adoption, and the steadily growing demographics that are young and tech-savvy. Such characteristics will be brought to light only through a proposed approach. A model specification developed in the light of these unique market characteristics, therefore, shall seek to assist electronic commerce businesses in Malaysia in current marketing, personalization of customer experiences, and operational efficiency improvement.

2. Literature Review

Key Theories of Consumer Behaviour in E-Commerce

E-commerce has significantly changed the means through which customers purchase their products, particularly in Malaysia, where online shopping has rapidly raised its demand after the COVID-19 pandemic. Understanding consumer behavior in the digital era is quite significant for e-commerce-based enterprises, especially when it turns to the habits of the online shopper. As can be seen from the standpoint of increasing income, it can offer an efficient marketing strategy. According to Beck & Ajzen (1991) Theory of Planned Behavior, attitudes, subjective norms, and perceived behavioral control influence online users in terms of their intention to carry out the behaviors online. This theory has been among the widely used theories to explain purchasing behavior, including online shopping. Decision tree algorithms can powerfully model TPB components by analyzing big datasets in search of patterns in attitudes, norms, and control factors that predict online shopping behavior. Opportunity of use, perceived ease of use, and attitude are other variables adopted in using segments of data, and decision trees create assumptions of consumers' decision rules for the final actual prediction of whether consumers are likely to make a purchase, hence; operating TPB in a usable, practical, and empirically testable manner.

Another old framework that this work embarks on is Howard Seth's model. This is a voluminous theoretical framework that claims to cover the consumer decision process. The model developed by John Howard and Jagdish Sheth is applied extensively in marketing and psychology to gain insight into how external variables influence choices that occur in an individual consumer(Howard & Sheth, 1968). It synthesizes different elements that include perceptual and learning constructs, to expand the view on the mechanisms of consumer decision-making. This model upholds the idea that consumer behavior is not only the result of immediate stimuli or overt acts but is an intricate interplay of many variables. It is therefore very significant to first understand this theory of behavior to form a base on which any prediction of consumer behavior stands, which will be necessary for coming up with an optimum e-commerce strategy. It is a high level of complexity that decision trees address by analyzing a group of factors together with the most determinable variable of influence on consumer decisions, such as marketing effort, consumer perception, and past purchase behaviors, when predicting future purchases for offering the best e-commerce optimization strategies.

Predictive Analytics in Modern E-Commerce

Predictive analytics has now become a powerful tool that helps in foreseeing the behaviors of customers in the digital world and provides noteworthy advantages to e-commerce businesses by making decisions based on their past data. The rationale for this has been related to the fact that consumer behavior is becoming increasingly complex in the digital space. Sales forecasting and decision-making based on sales have seen an enhancement in the Egyptian e-commerce industry with the help of predictive analytics, as found by Morsi (2020). This study identified that predictive analytics could reach the uncovered information about sales trends, optimize inventory, and enable better corporate planning. The results of this study indicate the role and importance of data-driven approaches in the highly competitive world of e-commerce in general but in developing countries like Egypt.

In Malaysia, e-commerce is growing at an unprecedented rate. According to Kathiarayan (2022), the COVID-19 pandemic has significantly increased e-commerce in Malaysia. In addition, data from the Malaysian Communications and Multimedia Commission shows a 48.8 percent increase in online shoppers from 2016 to 2018 (Alomari et al., 2021). A report by Statista (2024) predicts that this trend will continue, with 7.6 million more e-commerce users expected between 2024 and 2028. Due to the increase in online shopping, especially in Malaysia, there is a clear change in consumer behavior, which emphasizes the need to use decision trees for more thorough user data analysis. In the Malaysian context, a study by Choong et al. (2021) who focused on the importance of predictive analytics in predicting vegetable prices on e-commerce platforms, this study accurately addresses the unique challenges faced by farmers and the agricultural sector in Malaysia. Through surveys and experiments, this study identified that using advanced technology, especially in big data or machine learning can help solve the problem of food shortages. The study highlights that implementing an e-commerce platform with price forecasting tools can help SMI farmers better manage their products and avoid losses, emphasizing the need for such a platform in the Malaysian agricultural sector to achieve at least 80% accuracy in predicting vegetable prices, by that improves food supply management and supports government policy.

Furthermore, Bradlow et al. (2017) discuss how big data and predictive analytics are informing the functioning of retail organizations. The mentioned five elements, consisting of customer behavior, product knowledge, time, location, and sales channels, remain the focus of this study. These findings mean that, through the support of data sources, statistical tools, and theoretical concepts, merchants can get more accurate results in predicting customers' behavior, better inventory levels, and greater confidence in decision-making. It further iterates that effective analysis can only be realized through the integration of theory-driven methodologies with big data. It, therefore, calls for addressing ethical and privacy challenges in the business world.

Decision Trees and E-Commerce Behaviour

Among different predictive analytics techniques, decision tree algorithms gained popularity due to their simplicity and interpretability. A decision tree belongs to the class of nonparametric models that split the data into branches based on the values of selected features. Because of this, decision rules become straightforward to understand. Ketipov et al. (2023) made predictions of user behavior based on their personality attributes using decision trees and random forest models. The personality qualities studied here give a smaller picture, and the greater element of consumer behavior is left under-explored. Though this model shows excellent predictive ability, the personality trait focus of the study does present limitations. In this respect, future studies should be directed toward increasing the dimensions of user behavior to broaden the comprehensiveness and applicability of the findings. Overall, decision tree algorithms have shown promise in predicting customer behavior for online shopping, especially in the context of optimization or integrated approaches with other statistical methods. However, more research is necessary to understand their complete strength and weaknesses compared to other machine learning models. It is expected that the research will dwell on their performances when used independently.

This is further followed up by Ansari and Riasi (2019) in their study using the decision tree to find the J48 algorithm preference of shopping centers for higher-income households vis-a-vis the preference of stores for lower-income groups. Although these studies have contributed immensely to the understanding of demographics, their scope has been limited and focused on the site of shopping selection. The findings would be more in-depth and generalizable, furnishing a holistic insight into consumer behavior, by extending the scope of the study to include various acts of consumer behavior. For instance, Kareena and Kumar (2019) designed a hybrid classifier that involves decision trees and KNN utilized for user behavior prediction from Amazon reviews, and it outperformed the Naive Bayes approach. Although the hybrid model shows good predictive accuracy, it is based on a combination of algorithms that do not distinguish the individual performance strengths of each of these approaches. Future research should mainly focus on decision trees regarding independent assessment to establish best practices and gain insight into how effective the models are.

Zhenyu Liu (2019) adopted a stacking approach to combine decision tree models. The result was that he had high predictive performance for future purchases. Enabling this ensemble method to successfully take advantage of various models proves the value of such an approach. Otherwise, this success points out that only limited research effort was placed on investigating the performance of single decision trees. Further studies are therefore recommended to assess how well these decision trees perform separately and thus determine whether they work in isolation to be able to shed additional light on their performance taken on a stand-alone basis.

The research by Gkikas et al. (2022) used a hybrid model, combining decision trees with genetic algorithms to achieve an accuracy of more than 90% in consumer behavior forecasting. This is a strong improvement in the accuracy of the forecast, hence justifying the use of a hybrid model which uses genetic algorithms coupled with a decision tree. However, the emphasis that this study places on hybrid models belies the capabilities of standalone decision trees, and it suggests that future research is required to, in the first instance, investigate core decision tree models before considering enhancements through hybridization. Given this context, there is an apparent lack of understanding regarding how well standalone decision trees are doing, especially in the Malaysian e-commerce landscape. This paper, therefore, tries to fill that gap by assessing the effectiveness of decision trees in predicting consumer behavior based on demographics and psychographics with due regard to the ethical implications of deploying predictive analytics in e-commerce.

3. Methodology

The proposed methodology includes the following steps for the analysis of the consumer behavior data: online survey-based data collection, application of preprocessing techniques in datasets, decision tree machine learning algorithm for predictions after preprocessing, and classification metrics to compare algorithmic performances on all datasets, followed by test data analysis based on the parameters of the dataset. Each of these steps is explained in detail in the subsections describing the methodology.

Data Collection: This research performed data collection as an online survey among Malaysian consumers aged 18 and above who have been active in buying at least one item online using the internet. This approach will ensure comprehensive data is collected, which portrays a diversified range of consumer behaviors and their preferences. This data set consists of 85 features and 560 instances. This shows that this piece of data has a very comprehensive base for analysis. Described in Table 1 is the pre-processed data.

Table 1: Data Description before pre-processing

No	Attribute	Description
1	Gender	Consumer's gender
2	Age	Consumer's age
3	Level of Education	Consumer's level of education
4	Ethnicity	Consumer's ethnicity
5	Annual income	Consumer's annual income
6	Employment status	Consumer's employment status
7	Current residential location	Consumer's current residential location
8	Frequency of online shopping	How frequently the consumer visits online shopping platforms
9	Average time	The average time the consumer spends online shopping
10	Type of products	Products that are usually purchased by the consumer
11	Age	Consumer's age
12	Attitude (BA1, BA2, BA3, BA4, BA5, BA6, BA7)	Attitude factors influence online purchasing behaviour
13	Perceived Risk (BB1, BB2, BB3, BB4, BB5, BB6)	Perceived risk factors influence online purchasing behavior
14	Trust and Security (BC1, BC2, BC3, BC4, BC5, BC6)	Trust and security factors influence online purchasing behavior
15	Psychological	Psychological factors influence online purchasing

	(BD1, BD2, BD3, BD4, BD5, BD6	behavior
16	Hedonic Motivation (BE1, BE2, BE3, BE4, BE5)	Hedonic motivation factors influence online purchasing behavior
17	Promotion (BF1, BF2, BF3, BF4, BF5, BF6)	Promotion factors influence online purchasing behavior
18	Product Price (BG1, BG2, BG3, BG4, BG5, BG6)	Product price factors influence online purchasing behavior
19	Privacy (BH1, BH2, BH3, BH4, BH5, BH6)	Privacy factors influence online purchasing behavior
20	Emotional (BI1, BI2, BI3, BI4, BI5, BI6)	Emotional factors influence online purchasing behavior
21	Perceived Benefits (BJ1, BJ2, BJ3, BJ4, BJ5, BJ6)	Perceived benefit factors influence online purchasing behavior
22	Accessibility (BK1, BK2, BK3, BK4, BK5, BK6)	Accessible factors influence online purchasing behavior
23	Online Purchasing Behaviour (C1, C2, C3, C4, C5, C6, C7, C8, C9, C10)	Will the consumer purchase? (Output variable)

Pre-Processing: Generally, data pre-processing is an essential process for cleaning the data, which then ensures the reliability and consistency of the dataset. It began with data cleaning, by removing missing values and unnecessary columns, followed by a selection of the important attributes based on research needs. It also involved data transformation and application of SMOTE to handle class imbalance, thereby increasing the number of instances in the dataset from 560 to 1,126. The final data set consists of 30 features, and 1,126 instances, out of which there are five class labels: *"definitely wouldn't purchase," "probably wouldn't purchase," "might purchase," "probably would purchase," and "definitely purchase."*

Data Modeling: The pre-processed dataset after the cleaning process was then ready for machine learning algorithms to predict and analyze. Thereafter, 10-fold cross-validation was done to increase accuracy. Training and test samples were randomly selected from the base set for each classification, enabling models to be trained and tested and classification accuracy measures estimated for each classifier. Different methods of data mining were used in this study, but the focus remained on DT classifiers through the implementation of three methods, namely J48, Random Tree, and REPTree.

Decision Tree Algorithms: Decision trees are one of those flexible and interpretable machine learning algorithms that perform classification as well as regression. Examples of important decision tree algorithms in WEKA are J48, Random Tree, and REPTree. J48 can be seen as an implementation of C4.5; the construction of J48 is based on information gain to split the data and then prunes to avoid overfitting for the sake of interpretability. Random Tree constructs multiple unpruned trees using random subsets, which makes it more robust by ensemble learning. REPTree builds trees based on information gain and then prunes these with reduced error pruning, which is employed to guarantee efficiency combined with generalizability. These algorithms balance simplicity, robustness, and efficiency for a variety of data analysis tasks. Here's a bit more detail on each of the methods:

- **J48:** This is the implementation of the C4.5 algorithm that builds a decision tree by computation of information entropy. It is widely used since it is simple, besides being effective in generating understandable rules (Gyimah & Dake, 2019).
- **Random Tree:** Using the concept of several decision trees that consider random subsets of features and instances, this classifier provides stronger robustness, reducing the possibility of overfitting (Obiedat, 2020).
- **REPTree:** A fast decision tree learner. It builds a regression/decision tree using information gain/ variance reduction and prunes it using reduced-error pruning. It efficiently handles large datasets and noisy data (Abana, 2019).

Performance Measurement

The performance of the decision tree algorithms was evaluated in terms of accuracy, precision, recall, and F1 score. A complete analysis of the test data will be done by the different parameters of the dataset to meaningfully assess the performance of the models and point out improvements in the models. This would lead to the later extraction of useful understanding regarding how the performance, merits, and deficiencies of such models can be improved in the future.

- **Accuracy**: Measures the proportion of correct predictions made by a model out of the total number of predictions.
- **Precision**: Measures the proportion of positive class predictions that are correctly identified as positive. A higher precision score indicates that the model is more accurate and reliable in identifying positive samples.
- **Recall:** Measures a model's ability to correctly identify positive instances from the total number of actual positive instances in the dataset. It quantifies the proportion of true positive instances that the model successfully predicted as positive out of all the positive instances present.
- **F-measure:** Known as the F1-score, it combines precision and recall into a single metric, providing a balanced evaluation of the performance of a classification model, especially useful when dealing with imbalanced data sets in which the distribution of classes is irregular, and accuracy alone could be misleading.

5. Results and Discussion

This section presents some comparative experiments using some decision tree models, namely J48, Random Tree, and REPTree. These tests have been conducted on three different datasets. Cross-validation was used in the research investigation to evaluate the model. The metrics of assessment used in evaluating and comparing the classification models included accuracy, precision, recall, and F1 score. Results for each model run on each of the three different datasets are given in Table 2.

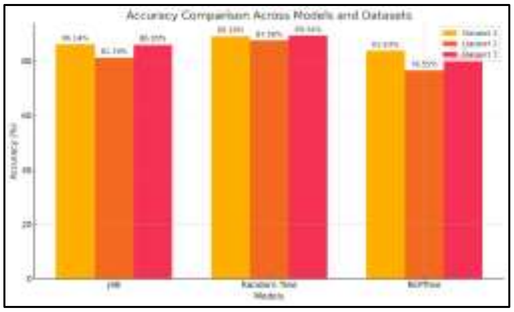
Table 2: Accuracy, Precision, Recall and F-Measure for Three Different Datasets

Dataset	Model	Accuracy	Precision	Recall	F-Measure
	J48	86.14%	0.859	0.857	0.858
1	Random Tree	89.16%	0.888	0.890	0.889
	REPTree	83.83%	0.830	0.836	0.833
	J48	81.26%	0.805	0.809	0.807
2	Random Tree	87.56%	0.870	0.872	0.871
	REPTree	76.55%	0.755	0.759	0.757
	J48	86.05%	0.857	0.856	0.857
3	Random Tree	89.34%	0.889	0.890	0.890
	REPTree	82.86%	0.820	0.825	0.823

All the datasets proved that the Random Tree model outperformed other models, as it had higher values of accuracy, precision, recall, and F1 scores, as observed in Table 2. While the J48 algorithm gave very performing and balanced values across all metrics, a bit more brittle, REPTree showed competitive results even if it came out as the weakest performance model compared to the other two. These findings are lent credence by the fact that the Random Tree model can handle several datasets and also that it is a working model for the classification tasks being investigated in this study. Because of this, the Random Tree model is not only reliable but also has a very high degree of precision; hence, it is suitable for application in further studies of consumer behavior and any other investigations that may be relevant to this area. The results of the study show the requirement in testing several models to come up with the best performance algorithms that could carry out certain data and serve desired goals.

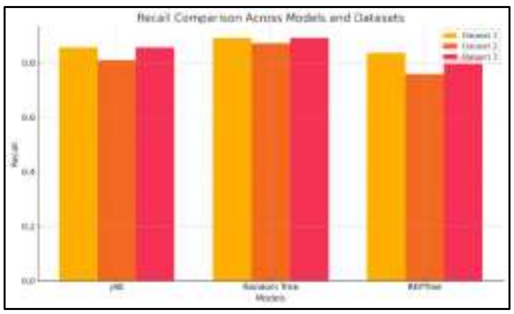
Figure 1: Performance comparison of Decision Tree Model Across Multiple Datasets: (a) Accuracy comparison across models and dataset; (b) F1 score comparison; (c) recall score comparison; (d) precision score comparison

a)



b)

c)



d)

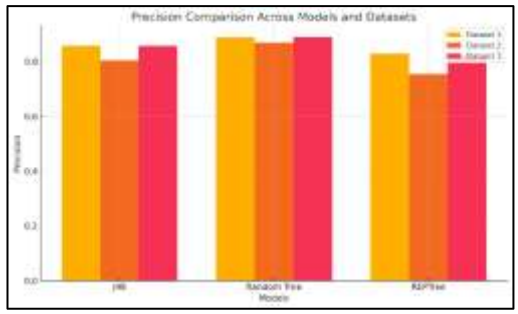


Figure 1 illustrates the detailed performance comparison across the datasets of the three Decision Tree models: J48, Random Tree, and REPTree based on four prime performance metrics, namely Accuracy, Precision, Recall, and F1 Score. From Figure 1(a), it can be seen that the F1 Score of the Random Tree model is higher across all data sets compared to the other two models, J48 and REPTree, indicating relatively better overall performance. Figure 1(b) illustrates the Recall of the models—their ability to retrieve all relevant instances. Also, like the F1 Score illustrated above, for all the datasets, the highest recall values correspond to the Random Tree model, which is thus more effective at retrieving actual positives.

Figure 1(c) shows the Precision, or in other words, the measure of goodness of models in terms of correct predictions among those forecasted as positive. Once again, the Random Tree model presents higher precision values for most of the datasets tested here, and thus, it is more precise when predicting positive cases. Finally, Figure 1(d) presents the Accuracy of the models, determined by the rate of correctly classified instances. Among the models, the highest accuracy is achieved by the Random Tree model, followed by J48 and then REPTree. The Random Tree model outperforms J48 and REPTree models on all performance measures in F1 Score, Recall, Precision, and Accuracy over the three datasets. That would mean that a Random Tree model would be most effective in terms of overall performance, therefore making it a better choice to predict consumer behavior in context.

5. Conclusion

In this present study, the J48, Random Tree, and REPTree decision tree models were explored in comparison to show a very good prediction of consumer purchasing behavior on e-commerce platforms in Malaysia. Models were evaluated based on accuracy, precision, recall, and F1 score using cross-validation. Surprisingly, the Random Tree model has sent consistent signals of its effectiveness concerning consumer behavior prediction. This model performed significantly better than the J48 and REPTree on all datasets tested. J48 was robust, presenting balanced metrics, but REPTree, though it received the lowest scores, also turned out competitive. The contribution of this research has added to the study and provided relevant information about the implementation of a decision tree model in e-commerce, specifically in Malaysia. These can serve as a guide for e-commerce firms in optimizing business marketing strategies and improving customer engagement through predictive analytics and machine learning. This provides the policy level of support for the need of businesses and policymakers to invest in advanced tools on data analytics that would drive evidence-based policymaking to enhance market competitiveness. Future research is needed into applying decision trees, as well as other machine learning models, to larger datasets to improve predictive accuracy and generalization. These models of machine learning, if applied appropriately, help the companies increase their operational efficiency enhance customer satisfaction and, in due course, scale up and surge forward in the highly competitive e-commerce domain.

Acknowledgment: The authors extend their sincere appreciation to Universiti Teknologi MARA, Cawangan Melaka, for the invaluable support provided in the completion of this article.

References

Abana, E. C. (2019). A decision tree approach for predicting student grades in a Research Project using Weka. *International Journal of Advanced Computer Science and Applications*, *10*(7), 285–289. https://doi.org/10.14569/ijacsa.2019.0100739

Alhakim, N. N., Stevia Septiani, & Eka Dasra Viana. (2023). Pengaruh Psychological Factors, Financial Literacy, dan Paylater Misuse terhadap Compulsive Buying pada Pengguna Aplikasi Paylater di Jabodetabek. *Jurnal Manajemen Dan Organisasi*, *14*(1), 57–68. https://doi.org/10.29244/jmo.v14i1.42798

Alomari, M. A., Ahmed Al-Qasem Al-Hadi, I., Yusoff, M. H., & Sulaiman, I. M. (2021). A Survey on Product Promotion via E-commerce Platforms - Case Study in Malaysia. *2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings*, *June*, 93–98. https://doi.org/10.1109/I2CACIS52118.2021.9495887

Ansari, A., & Riasi, A. (2019). Using decision trees to analyze the customers' shopping location preferences. *International Journal of Business Excellence*, *18*(2), 174–202. https://doi.org/10.1504/IJBEX.2019.099557

Beck, L., & Ajzen, I. (1991). Predicting dishonest actions using the theory of planned behavior. *Journal of Research in Personality*, *25*(3), 285–301. https://doi.org/10.1016/0092-6566(91)90021-H

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The Role of Big Data and Predictive Analytics in Retailing. *Journal of Retailing*, *93*(1), 79–95. https://doi.org/10.1016/j.jretai.2016.12.004

Castillo, M. J., & Taherdoost, H. (2023). The Impact of AI Technologies on E-Business. *Encyclopedia*, *3*(1), 107–121. https://doi.org/10.3390/encyclopedia3010009

Choong, K. Y., Raof, R. A. A., Sudin, S., & Ong, R. J. (2021). Time Series Analysis for Vegetable Price Forecasting in E-Commerce Platform: A Review. *Journal of Physics: Conference Series*, *1878*(1). https://doi.org/10.1088/1742-6596/1878/1/012071

Gkikas, D. C., Theodoridis, P. K., & Beligiannis, G. N. (2022). Enhanced Marketing Decision Making for Consumer Behaviour Classification Using Binary Decision Trees and a Genetic Algorithm Wrapper. *Informatics*, *9*(2). https://doi.org/10.3390/informatics9020045

Gyimah, E., & Dake, D. K. (2019). Using Decision Tree Classification Algorithm to Predict Learner Typologies for Project-Based Learning. *Proceedings - 2019 International Conference on Computing, Computational Modelling and Applications, ICCMA 2019*, 130–134. https://doi.org/10.1109/ICCMA.2019.00029

Howard, J. A., & Sheth, J. N. (1968). The theory of buyer behavior. *Journal of the American Statistical Association*, *65*(331), 1406–1407. https://doi.org/https://doi.org/10.2307/2284311

Kareena, & Kumar, R. (2019). A consumer behavior prediction method for e-commerce applications. *International Journal of Recent Technology and Engineering*, *8*(2 Special Issue 6), 983–988. https://doi.org/10.35940/ijrte.B1171.0782S619

Kathiarayan, V. (2022). The Impact of E-commerce Platforms in Malaysian Shopping During Covid-19 Pandemic. *Gyancity Journal of Electronics and Computer Science*, *7*(2), 2446–2918. https://doi.org/10.21058/gjecs.2022.72007

Ketipov, R., Angelova, V., Doukovska, L., & Schnalle, R. (2023). Predicting User Behavior in e-Commerce Using Machine Learning. *Cybernetics and Information Technologies*, *23*(3), 89–101. https://doi.org/10.2478/cait-2023-0026

Kiew, C. C., Abu Hasan, Z. R., & Abu Hasan, N. (2021). Factors Influencing Consumers in Using Shopee for Online

Purchase Intention in East Coast Malaysia. *Universiti Malaysia Terengganu Journal of Undergraduate Research*, *3*(1), 45–56. https://doi.org/10.46754/umtjur.2021.01.006

Kuswanto, H., Hadi Pratama, W. B., & Ahmad, I. S. (2020). Survey data on students' online shopping behavior: A focus on selected university students in Indonesia. *Data in Brief*, *29*. https://doi.org/10.1016/j.dib.2019.105073

Loh, C.-H., & Teoh, A.-P. (2021). The Adoption of Big Data Analytics Among Manufacturing Small and Medium Enterprises During Covid-19 Crisis in Malaysia. *Proceedings of the Ninth International Conference on Entrepreneurship and Business Management (ICEBM 2020)*, *174*(Icebm 2020), 95–100. https://doi.org/10.2991/aebmr.k.210507.015

Morsi, S. (2020). A Predictive Analytics Model for E-commerce Sales Transactions to Support Decision Making: A Case Study. *International Journal of Computer and Information Technology(2279-0764)*, *9*(1), 17–24. https://doi.org/10.24203/ijcit.v9i1.3

Obiedat, R. (2020). Developing decision tree-based models in combination with filter feature selection methods for direct marketing. *International Journal of Advanced Computer Science and Applications*, *11*(1), 650–659. https://doi.org/10.14569/ijacsa.2020.0110180

Pe, N., Gil-Saura, I., Rodríguez-Orejuela, A., & Ribamar Siqueira-Junior, J. (2020). *Purchase intention and purchase behavior online: A cross-cultural approach*. https://doi.org/10.1016/j.heliyon.2020.e04284

Sarangapani, S., Singh, S. S. R., Balasubramaniam, A., Hasan, W. S. W., Shaharuddin, N. binti, & Kanapathipillai, K. (2023). a Study of the Factors Influencing Purchasing Behaviour of Online Consumers in Klang Valley, Malaysia. *European Journal of Economic and Financial Research*, *7*(1), 120–150. https://doi.org/10.46827/ejefr.v7i1.1431

Statista. (n.d.). *Number of users of e-commerce in Malaysia 2017-2027*. Retrieved May 8, 2022, from https://www.statista.com/statistics/1351255/malaysia-number-of-e-commerce-users/

Vachkova, M., Ghouri, A., Ashour, H., Isa, N. B. M., & Barnes, G. (2023). Big data and predictive analytics and Malaysian micro-, small and medium businesses. *SN Business & Economics*, *3*(8), 1–28. https://doi.org/10.1007/s43546-023-00528-y

Zhenyu, Liu, X. M. (2019). Predictive Analysis of User Purchase Behavior Based on Machine Learning. *International Journal of Smart Business and Technology*, *7*(1), 45–56. https://doi.org/10.21742/ijsbt.2019.7.1.05