Research Proposal

Using Machine Learning for Customer Segmentation and Sales Prediction in E-Commerce Industries

Submitted by:

Contents

Introduction 1
1.1 Scope and Rationale 2
1.2 Study Aims 4
1.3 Objectives 4
1.4 Research Questions 4
Literature Review 5
Research Methodology 8
3.1 Research Philosophy 8
3.2 Research Approach 8
3.3 Data Collection 8
Data Analysis 10
4.1 Model Training 10
4.2 Performance Evaluation 11
4.3 Limitations 11
5. Ethical Consideration 12
6 Conclusion 13
Dissertation Overview and Headings 14
References 15
Appendix 19

Introduction

During recent years, artificial intelligence has seen a rise in practical applications in various industries such as healthcare, education, engineering, sales, entertainment, and transport (Bajaj et al. 2020). Machine learning, a branch of AI has also gained significant popularity in the marketing industry due to the vast amount of data that can be utilized to gain valuable insights that can drive important decisions for an organization's profitability and understanding of consumer behaviour.

Interestingly, this amount of data is expected to increase as more businesses use digital marketing to expand their operations (Boone et al. 2019). Forecasting sales allows businesses to gain meaningful insights for inventory management, budgeting, operational planning, and strategic decision making (Lau et al. 2018). It is an imperative step for strategic planning and making astute business decisions. As an e-commerce business becomes dependant on digital marketing platforms, using forecasting based on data analysis becomes crucial to maintain competitiveness (Cham et al., 2022)

Although traditional customer segmentation, which involves classifying a business's customer base into unique strata based on shared attributes and sales forecasting has contributed to successful business decisions in the past they are becoming increasingly obsolete (Turkmen 2022) in the present business context of digitalisation due to the rapidly changing demands of consumers. Such methods depend on a repository of past data combined with situational judgement decisions made by sales personnel along with analysing the current market trends. While these approaches yield results, they are essentially limited to the biasness of human input and interpretation, hence proving to be less efficient (Venkataramanan et al. 2024). Furthermore, these approaches may not account for variables such as changing consumer preferences, competitor activities, and differing economic conditions. Using machine learning to segment customers enables businesses to generate personalised profiles by analysing real time behaviour. These approaches adapt to changes in the market more effectively than traditional methods (Elhosseini 2023 et al. 2023).

The ongoing discourse by researchers such as Venkatraman have determined that while traditional approaches produce successful results when addressing factors such as evolving consumer demands, these methods have a limiting reliance on historical data that is formulated

or derived from market dependent decisions by stakeholders ,especially in the analysis of current market dynamics. However, Pandey and Elhosseini argue that the application of advanced analytical technologies such as machine learning techniques can provide a more nuanced approach to segmentation analysis of customers due to its capacity to intuitively learn and adapt in real time to changes in consumer behaviour. This debate extends and supports the notion that machine learning applications in ecommerce is a better approach from conventional methods. Additionally, there exists substantial research on the application of AI solutions in customer segmentation processes as well as sales prediction. While the entities seem divergent in studies by Kasem et al (2023) and (Cheriyan et al. 2022), there is an evident gap that necessitates the combination of both process/concepts in understanding the varying nature and impact of customers to businesses.

While there is substantial research on using artificial intelligence for customer segmentation and sales prediction, they exist as separate entities of discussion in most studies (V Kumar et al. 2018). Therefore, this gap needs to be addressed to understand the varying nature and importance of the customers. Consequently, using the identified segments of customers, e commerce businesses can predict their future purchase intentions and patterns (Baati and Mohsil 2020). Doing so will enable the businesses to focus their marketing efforts and relationship building measures with individuals that are highly likely contribute to the business through frequent purchases and being brand ambassadors to bring forth a positive reputation of the business in a challenging business environment (Khoa and Huynh 2023).

1.1 Scope and Rationale

As artificial intelligence techniques evolve, a variety of techniques such as clustering, classification, neural networks, decision trees and AdaBoost are employed to gain meaningful insights from data and leveraging it to make informed business decisions. The results of these techniques are utilised to enhance marketing efforts by businesses, leading to higher consumer engagement, and focused targeted marketing. The more concentrated marketing efforts businesses make, the better strategies they will be able to employ for achieving higher sales (Zulaikha et al. 2020)

Machine learning has the capability to process copious amounts of data efficiently and produce models that can be applied in the e-commerce industry (Sharda et al. 2018). In this context, customer segmentation and sales prediction are essential areas that benefit from machine learning. This study intends to integrate both approaches as it will enhance sales forecasting

and gain an understanding of consumer purchasing patterns, enabling businesses to make data driven decisions. The knowledge developed will bridge the gap between sales prediction and customer segmentation, offering practical understanding for enhancing marketing strategies, inventory management and tailored consumer experiences. Businesses, science, and technology sector are adopting the much-needed use of the machine learning techniques which has evolved into its reliance in fields such as manufacturing, healthcare, education, financial modelling, and marketing (Jordan and Mitchell 2020).

The scope of this research is confined to e-commerce due to the unprecedented challenges and opportunities online businesses present (Hagberg et al. 2016). Unlike retail businesses, online businesses often compete on a global scale and are not confined to a geographical area. Additionally, customer retention in online businesses requires continuous engagement, and personalised marketing.

Lastly, e-commerce businesses can use data analytics and machine learning to provide personalised purchasing experiences, thus enhancing customer satisfaction and brand loyalty (Zhang and Xiong 2024).

Therefore, the analysis investigates consumer behaviour in an online setting and the trends seen in sales made within this domain. Although the findings of the research will be suitable for online businesses only, the insights gained, and the methodology can be generalised to various other industries with similar patterns of data

The rationale for this research is derived from the growing necessity of e-commerce businesses to acquire advanced techniques that focus on utilising data in a systematic way. Traditional methods fail to keep track of the dynamic e-commerce business environment and rapidly changing consumer interests, competitor tactics, and extrinsic factors. Machine learning provides a powerful and meticulous alternative through the recognition of intricate patterns in datasets with vast number of values, that tradition methods cannot comprehend. Businesses can make focused decisions and marketing strategies based on accurate segmentation of customers, thus enhancing consumer engagement and brand loyalty. Additionally, reliable sales prediction models aid businesses achieve long term success though the optimization of operational planning, budgeting, and systematic inventory management. Although there is significant

research on customer segmentation and sales prediction individually, there exists a need to combine both aspects which is the purpose of this research.

1.2 Study Aims

As the competition among businesses rises in the digital space, it is crucial for businesses to leverage their data to reveal underlying trends and using the information revealed to modify and refine their marketing efforts. This paper aims to investigate the influence of machine learning in customer segmentation and sales prediction within the e-commerce industry. It also focuses on evaluating the overall efficiency of machine learning in enhancing the performance of a business

1.3 Objectives

- Determine which machine learning techniques are most effective to characterize e-commerce customer segments
- Evaluate the accuracy of the predictive sales model established using customer segmentation data in forecasting sales

1.4 Research Questions

- Which machine learning techniques are effective in customer segmentation and sales predictions?
- How do various data preprocessing procedures impact the performance of machine learning algorithms?
- What metrics are suitable for measuring the accuracy of the sales prediction model derived from customer segmentation data?
- What differences exist in the generated model's predictive accuracy between the various customer segments included in the e-commerce dataset?

Literature Review

Keywords	Scopus Results
"Machine Learning" AND "Customer Segmentation" AND "Sales Prediction"	3
"Supervised Learning" AND "Customer Segmentation" AND "Sales Prediction"	26
"Sales Prediction" AND "E-Commerce"	84
"Machine Learning" AND "Sales Prediction"	188
"Consumer Behaviour" AND "sales forecasting"	31
"Predictive Analytics" AND "Sales Models"	3
"Customer Segmentation" AND "Sales Prediction" AND "E-Commerce"	0
"Random Forest" OR "Decision Trees" AND "Sales Prediction"	68
"E-Commerce" AND "Machine Learning" not "Retail"	34
"Purchase History" AND "Customer Segments"	6

The relationship between machine learning and e-commerce has grown to be an area of increasing interest, particularly in view of customer segmentation and sales prediction. Though both subjects have been widely researched separately, there is a dire need to integrate them for optimal business results in the e-commerce industries.

Customer segmentation is a process of separating customers into distinct groups with identical behaviours and attributes for the purpose of focused marketing strategies. More conventionally,

the traditional ways of segmentation perform the task by considering demographic, geographic, or psychographic data. However, with the development of machine learning techniques, businesses can create dynamic insights in terms of consumer behaviour. As Christy et al. (2018) mentioned, segmentation is fundamental to both identifying customer needs and customising marketing effort. Their study extends the recency, frequency and monetary (RFM) analysis to include more sophisticated algorithms such as k-means clustering. It mentions that customer segmentation enhances the ability of businesses to address diverse customer groups more effectively. However, while so doing, it may limit them in traditional clustering methods and hence could suffer inability in addressing real-time market shifts.

All forecasting, from a sales viewpoint, has conventionally been done using various statistical methods that estimate future sales based on historical data. Although considered effective in several cases, such methods fail to capture the essence of the evolving e-commerce ecosystem, wherein customer behaviours are highly dynamic and driven by numerous external factors. Singh et al. (2020) mentions that random forest and gradient boosting are among the advanced models which perform efficiently in this domain. Nonlinear relationships can be represented, and volume and complex data is managed with greater predictiveness by these models. However, Bohanec (2017) highlights that while extremely useful and accurate, the black box machine learning models (whose internal working cannot be comprehended) pose a challenge due to their lack of interpretability.

Furthermore, different machine learning algorithms explored independently either for segmentation or for forecasting rarely combine both holistic approaches. Turkmen (2022) attempts to bridge this gap by incorporating k-means clustering into a statistical framework to segment customers and predict sales in an e-commerce setting. Although this model seems promising, it is limited by being based on a single clustering method. Exclusively depending on k-means clustering, when the complexity of customer behaviour is rising may lead to segmentation results that are overly simplified and cannot represent nuanced differences among customers, a concern similarly raised by Saxena et al. (2024). They go further into a variety of clustering techniques, including hierarchical clustering and DBSCAN, for a more robust framework of customer segmentation. However, as they indicate, such models are not very promising in larger datasets since the problem arises with visualization and interpretation. Another critical point of consideration is scalability and flexibility of the models adopted.

Raizada and Saini (2021) present the efficiency of random forest and extra tree regression on Walmart data for sales prediction, and they have received accuracy results above 98%. However, it does not take into consideration how these models would generalize across industries with different data structures and consumer behaviours. This remains a general problem in the literature: usually, the performance of algorithms is assessed in particular contexts without afterthoughts about their adaptiveness to general business environments.

Contrary to that, Liu et al. (2020) developed a more adaptive approach by proposing logistic regression combined with XGBoost for predicting customer repurchase behaviour. Their model overcomes an important weakness with traditional approaches since it oversees imbalanced datasets very well, a frequent problem in e-commerce, since a small percentage of customers drive a large percentage of sales. The study shows how machine learning could be applied not only to predict what customers will do in the future but also to develop better, targeted marketing initiatives, reinforcing the interdependence between customer segmentation and sales prediction.

lastly, Cheriyan et al. (2022) study different regression-based techniques for sales forecasting and report that random forest outperformed others, with an accuracy of 95.53%. Yet such technocratic success does not resolve the conceptual weakness of the exclusive use of regression-based techniques. As companies depend increasingly on real-time data, regression models, which rely on the stability of the relationships between the dependent and independent variables, cannot easily match the rhythm of the online commercial environment.

The literature review indeed shows that machine learning achieved promising results both in customer segmentation and sales forecasting, while the combination of the two fields has not been well explored. It is further observed that there is a tendency to be focused merely on methodological performance; providing accuracy rates or error metrics, while the critical issues of model interpretability, scalability, and cross-industry applicability are not considered. In this regard, future studies are needed to be more complete in modelling and integrating segmentation with prediction, exploiting the power of multiple algorithms to provide an even finer understanding of customer behaviour and sales trends.

Research Methodology

3.1 Research Philosophy

This study is based on scientific research using legitimate data, making realism the most suitable philosophical approach. It follows a quantitative framework, relying on deductive methods. The research does not consist of a hypothesis; however, it assumes a relationship between customer sales and the segment they belong to. Various statistical and mathematical tools will be used to analyse the data and draw conclusions from the hypotheses (Ishtiaq, 2019).

3.2 Research Approach

This study's methodology is based on deductive reasoning which is considered as the key component of a research approach based on positivism and objectivity. Deductive reasoning moves from a general problem to a specific conclusion, following a logical chronology of steps to determine whether a theory can be proven in particular circumstances (Saunders et al, 2019). Although there is no hypothesis for this thesis, deductive reasoning will apply as it involves a logical flow from prevalent rules to the specific case of the viability of machine learning to segment customers and use the aforementioned segments to predict future sales. However, a strict deductive approach may neglect an investigation of other methods that could even further disclose the use of machine learning for segmentation of customers and sales forecast.

The research will begin with the established principle that machine learning can manage an extensive amount of data and leverage it to gain meaningful insights. It will then continue with the proposition that machine learning techniques will use customer sales data to predict their future purchases and split the customers into different categories based on their purchase patterns. Lastly, based on the results of the previous steps, a discrete conclusion can be drawn over the efficacy of machine learning and the specific techniques employed, and their impact on a business' future planning.

3.3 Data Collection

This study will comprise of secondary data collected from University of California Irvine's machine learning repository website. It consists of an online retail dataset that portrays sales records of customers made from several countries. The dataset was chosen for this research

due to a considerable number of records amounting to more than 54,000, ensuring that the machine learning techniques used will have sufficient data to use, therefore ensuring a comprehensive analysis and insightful findings.

Data Analysis

4.1 Model Training

Once the data is prepared, it will run a series of machine learning methods for customer segmentation such as K-means clustering, which has the ability to separate customers into groups based on features such as purchasing behaviour, frequency of transaction and/or amount spent (Jain, 2010). It will be used because it is effective in grouping similar customers together according to their purchasing patterns. Hierarchal clustering may also be employed to build a ranking of clusters which will aid in understanding the relationship between customers at diverse levels. This approach allows an in depth understanding of customers by creating a hierarchy of clusters (Embrechts et al. 2013), which aids in the exploration of multi-level relationships among various consumer groups. Furthermore, this gives a very fine-grained understanding of the association between different tiers of customers, which is crucial while developing marketing strategies addressing various tiers of customers. A recency, frequency and monetary (RFM) model can also be used to determine how recently a customer made the purchase, how often they purchase and how much monetary amount they spent. This will assist in providing a data driven and systematic way to classify customers into segments pertaining to their purchasing patterns (Dogan et al. 2018). The incorporation of the RFM model brings in consistency regarding the segregation of customers. This model effectively quantifies customer value with direct relevance to their purchasing patterns, which is paramount when using data to drive marketing strategy. Through it, businesses can pinpoint the most valued customers and concentrate their resources effectively in predicting sales.

Once the customer data is segmented into clusters, the sales prediction algorithms will be executed. The chosen methods include linear regression which can predict sales using past purchasing patterns and customer features from the clustering techniques (Morsi 2020). This approach is chosen due to its ability to enhance data driven decision making and efficiently encapsulates the relationship between sales and the factors that influence them. Secondly, decision tree algorithm will be deployed to acquire the relationship between customer features and the sales made. Decision trees will be used as they are an effective means to capture complex non-linear data patterns in large data sets and provide a structured and interpretable model for decision-making (Mustakim et al. 2024). Since decision trees are interpretable, businesses can be shown how each customer feature influences sales and hence communicate findings with stakeholders easily.

Next, random forest regression method will be deployed to acquire to improve the prediction accuracy by aggregating the results of multiple decision trees. This method is expected to give a higher accuracy due to its ability to reduce overfitting, robustness, and ability to generalize (Naik et al. 2022). The capability of this method to avoid overfitting and generalize well to unseen data is essential in developing robust sales prediction models. A more accurate prediction will enable a business to strategize its marketing efforts more effectively and finally aid the research objective to leverage machine learning for business optimization.

4.2 Performance Evaluation

To evaluate the effectiveness of the sales prediction model, evaluation techniques mean absolute error (MAE) and root mean square error (RMSE) will be used. These metrics will measure the difference between the actual and predicted sales, therefore indicating the reliability and solidity of the machine learning model. If the values for RMSE and MAE are lower, it will indicate accurate performance. This will indicate that the predictions made by the model are closer to the actual sales figures.

4.3 Limitations

While the dataset used for this study provides a significant amount of data, it does contain limitations.

Missing or Incomplete Data: Since this study uses a secondary dataset retrieved from an online source, it is likely to contain missing or incomplete data which may negatively impact the accuracy of the machine learning results.

Geographical Constraints: While the dataset contains a considerable amount of data, it is confined to thirty-six countries, hence the results may not apply to other regions in the world.

Limited Historical Data: The dataset is likely to contain limited historical data, which may limit the model's capability to produce long term predictions or encapsulate the emerging behaviour trends of customers.

K Means Clustering: K means clustering requires the number of clusters to be preset. This can hinder the diversity of the customer segments and skew results. Despite this, it performs efficient computation and segmentation when the optimal number of clusters are identified.

The RFM model only has three variables (recency, frequency, and monetary value) which may cause the results of consumer behaviour techniques to be superficial. However, it is an

elementary process which provides a swift approach to identify high-value customers, thus making it an appropriate choice for initial segmentation.

Linear regression: This approach operates on assumptions of the relationships among variables as being linear, which may not be accurate for the dataset being used as it contains a wide array of rows with thousands of data entries. Despite this, it is a useful method as it is simple to interpret, simplifying the understanding of the result of independent variables on sales. Consequently, this can result in valuable insights for e-commerce decision making.

5. Ethical Consideration

The secondary data used in this research is available online on University of California Irvine's machine learning repository and is open source. It is therefore available for research purposes and using this data does not infringe the privacy of any individuals whose sales data is recorded.

Conclusion

The research emphasizes the increasing value of machine learning in e-commerce businesses. It illustrates the process of identifying customers and using their purchasing patterns to forecast the future demand, facilitating businesses to make strategic sales and marketing decisions. As traditional customer segmentation and sales prediction methods become obsolete in the rapidly evolving world, machine learning offers a robust and data driven approach to this problem. Techniques such as hierarchical clustering and k means clustering can provide a reliable mechanism for segmenting customers into distinct groups that organisations can cater to according to their unique needs.

The significance of this study lies in its potential to transform the approach e-commerce businesses take to understand their customers and predict future sales, resulting in numerous significant contributions.

By combining customer segmentation and sales prediction, e-commerce businesses can gain a deeper understanding of their consumers and their purchasing behaviour. Secondly, the advanced segmentation approaches in this study can enable businesses to create personalised customer experiences. This can lead to higher customer satisfaction and brand loyalty. Also, by incorporating the two methods, this research has the capability to enhance the forecast accuracy of sales.

Another crucial impact of this study is that the findings can be used in direct applications of the e-commerce industry, which may result in the conception of exclusive strategies and tools for customer relationship management and optimisation of sales. Lastly, the methods proposed in this study can be applied to various e-commerce sectors and scaled to assist the expanding convolution of online retail conditions.

Dissertation Overview and Headings

Introduction
1.1 Background and significance
1.2 Research Objectives
1.3 Research questions
1.4 Scope and Limitations
Literature Review
2.1 Machine learning for e-commerce: Concepts and challenges
2.2 Customer Segmentation and Sales Prediction
2.3 Intersection of customer segmentation and sales prediction in machine learning
2.4 prior studies on Machine Learning for customer segmentation and Sales Prediction
2.5 Gaps in current literature
Research Methodology
3.1 Research Approach
3.2 Data Collection
3.3 Data Analysis Techniques
3.4 Ethical Consideration
Findings and Analysis
4.1 Analysis of Findings
4.2 Machine Learning Analysis
Discussion
5.1 Integration of Findings
5.2 Implications of Research
5.3 Application in Future Research
Conclusion
6.1 Synopsis of Findings
6.2 Contribution of Research
6.3 Recommendations
6.4 Conclusion

References

Baati, K. and Mohsil, M. (2020) Real-Time Prediction of Online Shoppers' Purchasing Intention Using Random Forest. *IFIP Advances in Information and Communication Technology* 43–51.

Bajaj, P., Ray, R., Shedge, S., Vidhate, S. and Shardoor, N. (n.d.) *SALES PREDICTION USING MACHINE LEARNING ALGORITHMS*. *International Research Journal of Engineering and Technology* .

Banerji, R. and Singh, A. (2024) Do social media marketing activities promote customer loyalty? A study on the e-commerce industry. *LBS Journal of Management & Research* 22 (1), Emerald93–109.

Bharadwaj, A., El Sawy, O.A., Pavlou, P.A. and Venkatraman, N. (2013) Digital Business Strategy: Toward a Next Generation of Insights. *MIS Quarterly* 37 (2), 471–482.

Bohanec, M., Kljajić Borštnar, M. and Robnik-Šikonja, M. (2017) Explaining machine learning models in sales predictions. *Expert Systems with Applications* 71, 416–428.

Boone, T., Ganeshan, R., Jain, A. and Sanders, N.R. (2019) Forecasting sales in the supply chain: Consumer analytics in the big data era. *International Journal of Forecasting* 35 (1), 170–180.

Burrell, G. and Morgan, G. (2017) Sociological paradigms and organisational analysis: Elements of the sociology of corporate life. Routledge.

Cham, T.-H., Cheah, J.-H., Memon, M.A., Fam, K.-S. and László, J. (2022) Digitalization and its impact on contemporary marketing strategies and practices. *Journal of Marketing Analytics* 10 (1),.

Cheriyan, S., Ibrahim, S., Mohanan, S. and Treesa, S. (2022) *Intelligent Sales Prediction Using Machine Learning Techniques*. 278–286.

Christy, A.J., Umamakeswari, A., Priyatharsini, L. and Neyaa, A. (2018) RFM Ranking – an Effective Approach to Customer Segmentation. *Journal of King Saud University - Computer and Information Sciences* 33 (10)

Dogan, O., Aycin, E. and Bulut, Z. (2018) Customer Segmentation by Using Rfm Model and Clustering methods: a Case Study in Retail Industry. *International Journal of Contemporary Economics and Administrative Sciences* 8, 1–19.

Dullaghan, C. and Rozaki, E. (2017) Integration of Machine Learning Techniques to Evaluate Dynamic Customer Segmentation Analysis for Mobile Customers. *International Journal of Data Mining & Knowledge Management Process* 7 (1), 13–24.

Embrechts, M.J., Gatti, C.J., Linton, J. and Roysam, B. (2013) Hierarchical Clustering for Large Data Sets. *Advances in Intelligent Signal Processing and Data Mining* 197–233.

Hagberg, J., Sundstrom, M. and Egels-Zandén, N. (2016) The digitalization of retailing: an exploratory framework. *International Journal of Retail & Distribution Management* 44 (7), Emerald694–712.

Heron, J. (1996) Co-operative inquiry: Research into the human condition. Sage.

Jafari, R. (2022) Hands-on data preprocessing in Python: Learn how to effectively prepare data for successful data analytics: Packt Publishing.

Jain, A.K. (2010) Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31 (8), 651–666.

Jordan, M.I. and Mitchell, T.M. (2020) Machine learning: Trends, perspectives, and prospects. *Science* 349 (6245), 255–260.

Kansal, T., Bahuguna, S., Singh, V. and Choudhury, T. (2018) Customer Segmentation Using K-means Clustering. *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)* 135–139.

Khoa, B. and Huynh, T. (2023) The influence of social media marketing activities on customer loyalty: A study of e-commerce industry. *International Journal of Data and Network Science* 7 (1), 175–184.

Kilani, M.A. and Kobziev, V. (2016) An Overview of Research Methodology in Information System (IS). *Open Access Library Journal* 03 (11), 1–9.

Lau, R.Y.K., Zhang, W. and Xu, W. (2018) Parallel Aspect-Oriented Sentiment Analysis for Sales Forecasting with Big Data. *Production & Operations Management* 27 (10), 1775–1794.

Liu, C.-J., Huang, T.-S., Ho, P.-T., Huang, J.-C. and Hsieh, C.-T. Z Lv (editor), (2020) Machine learning-based e-commerce platform repurchase customer prediction model. *PLOS ONE* 15 (12), e0243105.

Mahesh, B. (2018) Machine Learning Algorithms -A Review. *International Journal of Science and Research (IJSR) ResearchGate Impact Factor* 9 (1).

Mahmoud SalahEldin Kasem, Hamada, M. and Islam Taj-Eddin (2023) Customer profiling, segmentation, and sales prediction using AI in direct marketing. *Neural Computing and Applications* 36, Springer Science+Business Media.

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)

Mustakim, N.A., Abdul Aziz, M. and Abdul Rahman, S. (2024) Predicting Consumer Behaviour in E-Commerce Using Decision Tree: A Case Study in Malaysia. *Information Management and Business Review* 16 (3(I)), 201–209.

Naik, H., Yashwanth, K., P, S. and Jayapandian, N. (2022) *Machine Learning based Food Sales Prediction using Random Forest Regression*. 998–1004https://ieeexplore.ieee.org/abstract/document/10009277 Accessed.

Raizada, S. and Saini, J.R. (2021) Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting. *International Journal of Advanced Computer Science and Applications* 12 (11).

Rokach, L. and Maimon, O. (2005) Clustering Methods. *Data Mining and Knowledge Discovery Handbook* 321–352.

Saunders, M., Lewis, P. and Thornhill, A. (2019) *Research Methods for Business Students*. 8th Edition. Harlow: Pearson.

Saxena, A., Agarwal, A., Binay Kumar Pandey and Pandey, D. (2024) Examination of the Criticality of Customer Segmentation Using Unsupervised Learning Methods. *Circular Economy and Sustainability* 4, Springer Nature.

Sharda, R., Delen, D. and Turban, E. (2018) *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*. 4th Edition. Boston: Pearson Education, Inc150–165.

Turkmen, B. (2022) Customer Segmentation with Machine Learning for Online Retail Industry. *The European Journal of Social & Behavioural Sciences* 31 (2), 111–136.

V Kumar, Reinartz, W. and Springer-Verlag Gmbh (2018) *Customer Relationship Management Concept, Strategy, and Tools*. Berlin Springer Berlin Springer.

Venkataramanan, S., Sadhu, A.K.R., Gudala, L. and Reddy, A.K. (2024) Leveraging Artificial Intelligence for Enhanced Sales Forecasting Accuracy: A Review of AI-Driven Techniques and Practical Applications in Customer Relationship Management Systems. *Australian Journal of Machine Learning Research & Applications* 4 (1), 267–287.

Zhang, Q. and Xiong, Y. (2024) Harnessing AI potential in E-Commerce: improving user engagement and sales through deep learning-based product recommendations. *Current Psychology* Springer Science and Business Media LLC.

Zulaikha, S., Mohamed, H., Kurniawati, M., Rusgianto, S. and Rusmita, S.A. (2020) Customer Predictive Analytics Using Artificial Intelligence. *The Singapore Economic Review* 1–12.

Ishtiaq, M. (2019) Book Review Creswell, JW (2014). Research Design: Qualitative, Quantitative and Mixed Methods Approaches. Thousand Oaks, CA: Sage. English Language Teaching 12 (5), 40.

Appendix

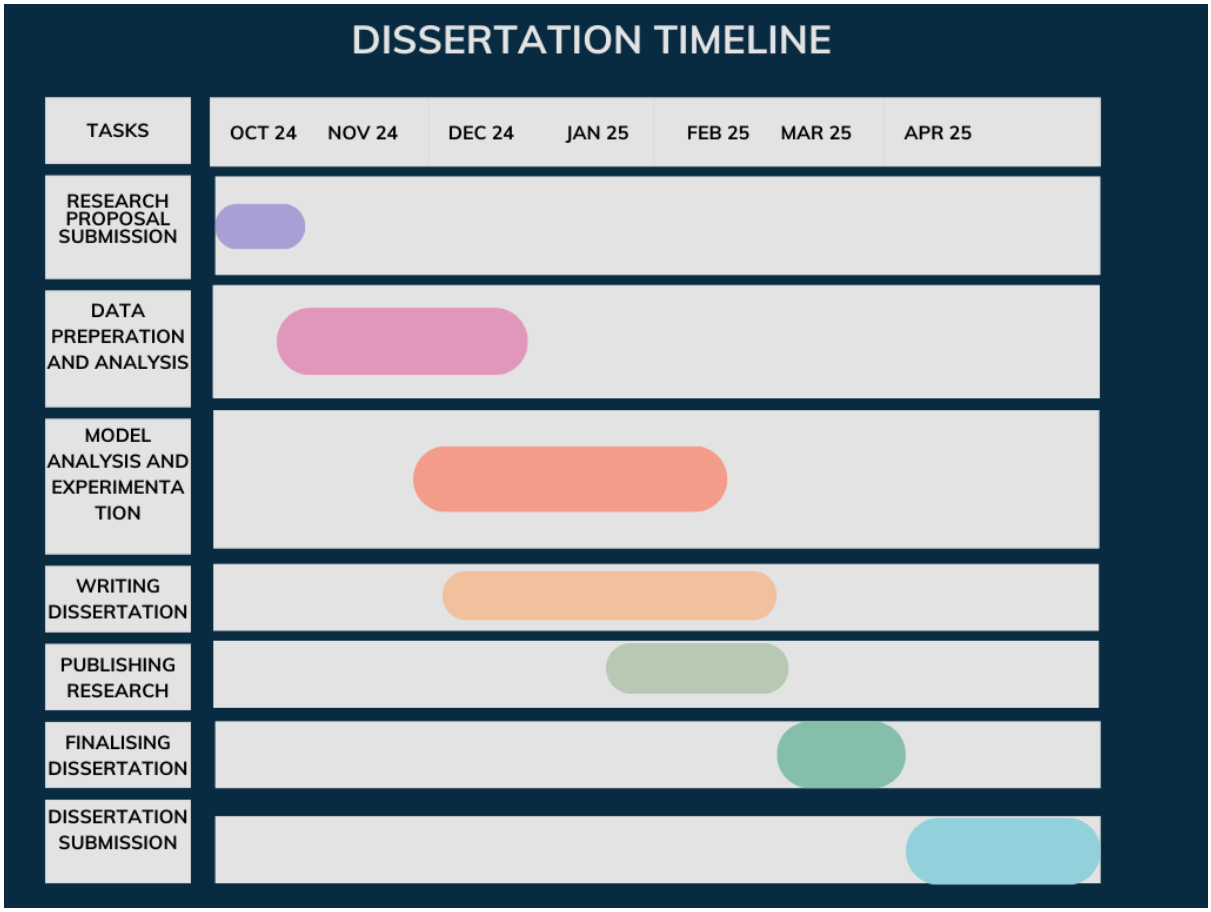

Figure 1: Dissertation |Timeline