Understanding Consumer Behavior in the Retail Sector Using RFM Segmentation and Machine Learning: An Analysis

Article · August 2024

DOI: 10.52783/eel.v14i3.1783

Dr.Mallikarjuna Reddy Doodipala
GITAM Deemed to be University

Janardhana Rao Nidamaluri
QIS College of Engineering and Technology

Understanding Consumer Behavior in the Retail Sector Using RFM Segmentation and Machine Learning: An Analysis

***Dr. B. Rajeshwari, **Dr Mallikarjuna Reddy Doodipala, ***Dr.K.Kiran Kumar Varma, ****Dr. P. Pattabhi Ram, ****Dr. Janardhana Rao N**

*Asst. Professor, G.Narayanamma Institute of Technology & Science (For Women), Hyderabad, India.

**Associate Professor Department of Mathematics, gitam Hyderabad

***Associate Professor, Dept of EM&H, SRKR Engineering College, Bhimavaram.

****Professor & HO, Narasaopeta Engineering Colleges (Autonomous), Narasaraopet.

*****Associate professor and HoD, Department of Master of Business Administration, QIS College of Engineering and Technology, Autonomous, Ongole

Abstract

The current study uses advanced analytics to examine the intricate dynamics of consumer behavior in the Indian retail sector. We created a preprocessed retail dataset with 495,478 retail clients in it. The study aims to forecast consumer behavior, the study uses a variety of techniques, such as temporal analysis, Box-Cox transformation, and a marketing analysis method called RFM (Recency, Frequency, Monetary) segmentation. Additionally, it makes use of a variety of supervised machine learning models, such as RandomForest, AdaBoost, ExtraTrees, LGBM, and XGBoost, among which ET and XGBClassifier have demonstrated the highest levels of accuracy in customer lifetime value predicting. The study stated that the machine learning models' performance measures are remarkable: 92.40% accuracy, 92.27% precision, 92.40% recall, 92.28% F1 score, and 97.39 AUC. The study's findings validated the durability of machine learning techniques and demonstrated the model's accuracy in predicting customer lifetime value clusters. Important conclusions from RFM analysis show that it has a special value in offering important insights into customers and their behavior. This study sets a new benchmark for retail analytics by providing a scalable and effective technique for research projects in the future that use data analytics to understand and forecast customer behavior across various business entities.

Keywords: *Consumer Behavior Analytics, Supervised Machine Learning, Retail Analytics, RFM method*

1. Introduction

The RFM (Recency, Frequency, Monetary) analysis technique is a straightforward yet highly effective method for understanding and evaluating consumer behavior based on purchase history. It quantitatively categorizes and classifies customers based on the recency, frequency, and monetary value of their most recent transactions. The primary goal is to identify and target the most valuable customers to execute focused and precision-targeted marketing campaigns. While RFM analysis provides a quantitative method to categorize and classify customers based on transactional behavior, it is important to recognize that, like any model, it may incorporate biases related to demographic characteristics not included in the analysis and the methods used for data collection. Therefore, while RFM offers a robust framework, its outcomes must be interpreted with an understanding of these limitations. This technique is based on the widely recognized marketing principle that "80% of your business is generated by 20% of your customers." RFM is a strategic methodology employed to analyze and estimate the value of a client based on three crucial data points: recency, frequency, and monetary value. The recency metric indicates the timing of the consumer's most recent transaction, frequency measures how often the consumer makes purchases, and monetary value represents the amount of money spent by the customer.

RFM analysis is a beneficial approach that can provide essential insights into clients and their behavior. However, it is important to acknowledge that this strategy does not consider other critical factors influencing the customer experience. To achieve superior outcomes, more comprehensive marketing techniques might include additional elements, such as the type of item purchased or client campaign responses. Furthermore, RFM analysis does not account for client demographics, such as age, gender, and ethnicity. Thus, marketers must employ a more extensive and sophisticated strategy that considers a broader array of aspects to gain a more precise and practical understanding of customers. The literature offers numerous methods for data integration that can be useful in this context.

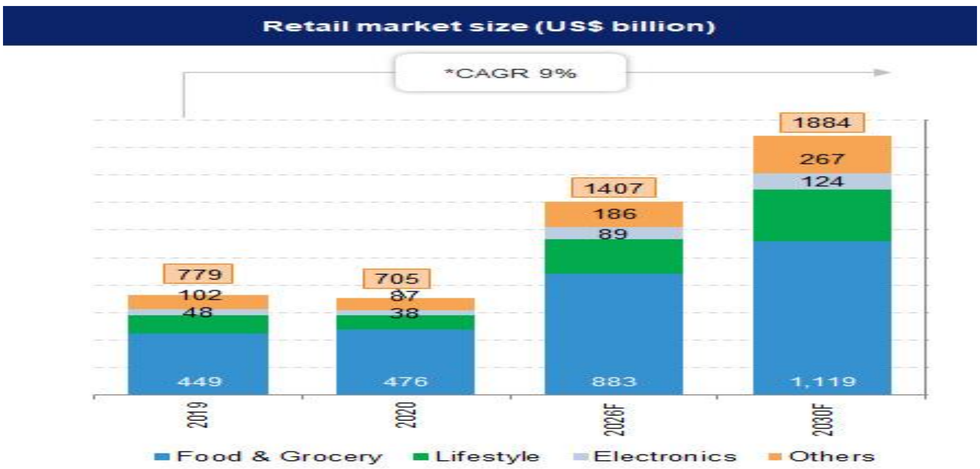
While the RFM method is powerful, it currently excludes direct consideration of customer demographics, which are pivotal in understanding consumer patterns. Future extensions of this work will aim to integrate these demographic factors, potentially enhancing the predictive capacity of our models and offering a more comprehensive picture of consumer behavior. Moreover, it is important to acknowledge that RFM relies exclusively on historical customer data, suggesting that it may not reliably predict future customer behaviors. Predictive approaches, on the other hand, can reveal potential client behavioral trends that may go unnoticed by RFM analysis. This indicates that while RFM is useful for studying customer data, its limitations in forecasting future behavior necessitate the integration of more sophisticated predictive methodologies.

In the context of the current research study, RFM values can be easily computed and understood; however, they represent only a single aspect of client behavior. To develop accurate prediction models, data analysts utilize diverse data sets encompassing client needs, opinions, socio-economic traits, and relationship data, among other factors. Collecting such data can be challenging, as small and mid-sized organizations often lack systematic methods for data gathering. Calculating RFM scores for practical use requires specialist analytical abilities or advanced mathematical proficiency. Furthermore, like any other model, the complexity of RFM models can range from basic to advanced. The RFM segmentation process begins by categorizing clients based on their recency score, frequency score, and monetary score. By employing an RFM scoring system, one can develop a successful marketing plan by generating client RFM segments.

Within the existing literature, several methods have been proposed for customer categorization. In this study, it utilize the Box-Cox transformation to preprocess the data. This technique, which has not been previously employed in similar research, is used to ensure that the data conforms to a normal distribution. Additionally, this study represents the first investigation to apply client segmentation specifically within the United Kingdom. The research involves a comparison of various techniques.

1.1 Indian retail market

India's retail market is expected to rise at a rate of nine percent between 2019 and 2030, from US$ 779 billion in 2019 to US$ 1,407 billion by 2026, and more than US$ 1.8 trillion by 2030, according to Kearney Research. By the end of 2025, the direct selling market in India is projected to be worth US$7.77 billion. In spite of hitherto unseen obstacles, the Indian consumer story remains strong.India is the country with the third-highest number of online shoppers, after the US and China. By 2030, it is anticipated that the leading logistics companies of today would transport 2.5 billion Direct-to-Consumer (D2C) goods. Over the next ten years, there projected a nine-fold growth in the penetration of online used automobile transactions.Recent industry reports state that the e-commerce sector had an astounding 36.8% YoY surge in order volumes. The tendency of consumers to purchase online year-round indicates the maturity attained by Indian e-commerce firms. This inclination is rapidly shifting. There were 7.8 billion e-commerce transactions per day as of December 2022. India's e-commerce population is predicted to grow from +150 million in 2020 to about 500 million by 2030. By 2030, the digital economy in India is projected to reach US$ 800 billion, while the GMV of the e-commerce sector is projected to reach US$ 350 billion.



Source: IBEF Report

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 3 (2024)
http://eelet.org.uk

1.2 Objectives of the Study

- To examine customer behavior and forecast Customer Lifetime Value (CLTV) using a retail dataset.
- To gain insights into the purchase behaviors of various client segments by analyzing their Recency, Frequency, and Monetary (RFM) values.
- To utilize machine learning methods to predict CLTV, providing valuable information for targeted marketing strategies and customer relationship management.
- To enhance customer interaction and optimize resource allocation by understanding and forecasting client behaviors

1.3 Research Questions

The study identify the critical inquiries that guide the direction and focus of our investigation. These questions are meticulously crafted to examine essential elements of customer behavior research and to uncover the patterns and factors that influence effective business strategies.

- What are the most effective methods for segmenting clients based on their purchasing activity, specifically in terms of recency, frequency, and monetary value of their purchases?
- What are the expenditure trends and financial contributions of customers within a designated six-month timeframe?
- Which machine learning models exhibit superior performance in forecasting customer lifetime value, and what are their primary performance metrics?

Customer churn prediction involves applying machine learning algorithms to identify consumers who are most likely to terminate their relationship with a company. This process typically entails training a supervised machine learning model using historical data, such as consumer purchase history, account activity, and customer support interactions. The trained model is then used to generate predictions with new data, including the likelihood of a customer discontinuing service within a specified period. Customer churn prediction is a classification task, where the machine learning model is utilized to determine whether a customer is likely to churn.

2. RELATED WORK

Recent literature has increasingly focused on integrating supervised machine learning (ML) models with traditional statistical models to better understand the decision-making processes of various stakeholders, such as consumers, producers, learners, and farm managers [16, 18, 26]. Existing studies have examined farmers' adoption decision behaviors by analyzing primary survey data, which includes factors related to producer perception, adoption decisions, and economic and socio-demographic issues [21-25]. Evaluating the performance of different ML models for prediction and forecasting purposes has proven to be valuable [19, 20, 27].

Research [10, 28] indicates that enhancing marketing performance necessitates the combined use of customer segmentation and consumer targeting. These interrelated objectives are integrated into a systematic methodology, though they face the challenge of achieving unified optimization. To address this issue, the authors propose employing the K-Classifiers Segmentation approach. This method utilizes transactional data from a large number of consumers to yield more precise clustering results. The authors also highlight that identifying an optimal segmentation approach is inherently NP-hard, necessitating the development of numerous carefully designed suboptimal clustering algorithms. Their analysis of consumer segments generated through direct grouping methods revealed significantly better outcomes compared to traditional statistical approaches.

The research has introduced a novel clustering methodology that shares similarities with both the K-means and K-medoids algorithms [11, 29-30]. These algorithms are recognized as partitional methodologies. However, the newly proposed technique does not guarantee an optimal solution in all cases, although it does reduce cluster error criteria. The study [12] has proposed a personalized recommendation system designed to cater to individual user preferences and requirements. This system employs weighted frequent pattern mining to identify the most frequently occurring patterns. The authors conducted consumer profiling using the RFM model, a well-established method for identifying new clients. The proposed system assigns diverse weights to each transaction to generate association rules from the mining process. Implementing the RFM model enhances the accuracy of client recommendations, which in turn can lead to increased profitability for the company. The researchers performed a comprehensive analysis focused on predicting client attrition. They utilized logistic regression and successfully extracted transactional data to develop a novel prediction model [13].

Their experiments demonstrated that personalized marketing techniques could help identify and retain clients with a high likelihood of leaving.

Additionally, analysts have proposed a complex, multi-dimensional approach aimed at improving customer lifetime value, increasing customer satisfaction, and positively influencing consumer behavior [14]. Through thorough analysis, it was noted that consumers are a diverse group with unique wants, interests, and aspirations. To address this diversity, segmentation approaches can accurately identify and understand customer needs and expectations, thereby facilitating the provision of exceptional services to key customers. The authors developed a unique strategy that combines RFM and lifetime value methodologies for segmentation. This approach was executed in two distinct phases [15]: the first phase involved statistical analysis, while the second phase used clustering. The primary objective was to apply K-means clustering in a two-phase model and subsequently use a neural network to refine the segmentation process.

2. Research Methodology

The methodology employed in our study involves an integrated approach that begins with thorough data preprocessing, which includes cleaning, normalization, and transformation. A key component of our analysis is feature engineering, where we extract essential metrics such as RFM values from client transaction data. By employing a combination of supervised and unsupervised learning techniques, we use RFM analysis to segment customers and apply various machine learning models to forecast customer lifetime value. This comprehensive methodology is designed to provide a deep understanding of customer behavior and support predictive analytics.

Dataset Description: Our study focuses on data from the United Kingdom, as detailed in Table It retrieved all observations related to internet shopping within this specific country. During the preprocessing phase, we examined the presence of null values. In relational databases, a null value indicates that the value in a particular column is either unknown or missing. It is important to differentiate between a null value, an empty string (often encountered in character or datetime data types), and a zero value (typically found in numeric data types). These distinctions are crucial for accurate data interpretation and manipulation. In our dataset for the United Kingdom, there are 133,600 instances of missing or unknown values, which represents approximately 26.97% of the total sample size of 495,478.

Table 1. Dataset Attributes & Description

Attribute Name	Description
Invoice Number	A unique identifier, consisting of 6 digits, assigned to each transaction
Stock Code	A distinct 5-digit identifier is allocated toeach product
Quantity	The number of units of the product involvedin each transaction
Invoice Date	The date and time when the invoice wasgenerated
Unit Price	The cost of a single unit of the product
Customer ID	A unique 5-digit number is assigned to eachcustomer
Country	The name of the country associated with thetransaction

RFM values are computed as follows: Recency is calculated by determining the number of days between the most recent invoice date and the maximum date for each customer. Frequency is assessed by counting the number of occurrences of each customer ID. Monetary value is computed by multiplying the quantity of items by their unit price to find the total price for each customer. The monetary value for each client is then determined by summing the total prices across all transactions up to the specified end date.

Data Preprocessing: The data cleansing process comprised several critical stages. Initially, we identified and addressed missing values by either removing or imputing them. Special attention was given to anomalies, such as negative revenue figures, which could indicate returns or data entry errors. Any such anomalies were either corrected or excluded to ensure the quality and reliability of the data.

A crucial aspect of the data preparation phase involved converting the data into an appropriate format for analysis. This included:

- Transforming the 'InvoiceDate' into a datetime format to facilitate time-series analysis and enable the extraction of data elements such as year and month.
- Focusing on specific subsets of the data, such as transactions from particular countries (e.g., the United Kingdom) or defined periods (e.g., a six-month window), to enhance the relevance and manageability of the study.

Feature Engineering

Feature engineering was pivotal in enhancing the dataset's suitability for predictive modeling:

- RFM Attributes: The development of Recency, Frequency, and Monetary value metrics was essential. Recency measures the time elapsed since the most recent purchase, Frequency quantifies the number of transactions, and Monetary value represents the total expenditure by each customer.
- Temporal Attributes: To facilitate a detailed analysis over a six-month period, new features were created to accurately capture recent consumer activity, including the total revenue generated within this timeframe.
- Binary & Categorical Attributes: Converting categorical data, such as customer segments, into binary formats was necessary for integration into machine learning models.

Analytical Modeling: This section delves into a diverse array of statistical and machine learning methodologies aimed at uncovering the intricacies of customer behavior and predicting future purchasing patterns. The chapter focuses on leveraging our meticulously curated dataset through both traditional and advanced analytical techniques.

- ***RFM Analytics***: RFM Analysis is a foundational method that segments customers based on their historical purchase behavior. It categorizes customers into distinct groups according to Recency, Frequency, and Monetary values.
- ***K-Means Clustering (Unsupervised Learning):*** Further segmentation of clients is achieved by integrating RFM values with specific behavioral patterns, such as spending over the past six months, through K-Means clustering.
- ***Binary Classification (Supervised Learning):*** Various models, including Random Forest, AdaBoost, ExtraTrees, LGBM, and XGBoost, were employed to predict customer lifetime value segments. These models were selected for their ability to handle large datasets and their effectiveness in classification tasks.

4. Experiments and Results

This section presents a thorough analysis of our empirical research, focusing on the four primary research questions previously outlined. Using Python for our analytical procedures, we extensively explore the retail dataset, with a particular emphasis on the purchasing behaviors of Indian customers. This section details each phase of our experimental approach, including data processing and the application of various analytical models. It concludes with a comprehensive analysis and discussion of the findings and insights derived from these investigations. The objective of this study is twofold: to address the research questions and to demonstrate the practical application of advanced data analysis techniques in understanding complex customer behaviors.

RFM Segmentation

The RFM segmentation approach utilized in this study involves classifying clients based on three key metrics: Recency (the time elapsed since a customer's most recent purchase), Frequency (the frequency of their purchases), and Monetary Value (the amount they spend). The dataset undergoes initial cleansing and preprocessing, which includes addressing missing values and converting the 'InvoiceDate' field into an appropriate datetime format. Recency is calculated by subtracting the date of the most recent purchase from the current date. Frequency is determined by counting the number of transactions per customer, while Monetary Value reflects the total expenditure per customer.

RFM segmentation provides a comprehensive view of the customer base, considering multiple dimensions of purchasing behavior. Customers are categorized into various segments, such as high-value customers who make frequent and recent purchases, or low-value customers who have infrequent transactions and minimal spending. This segmentation

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 3 (2024)
http://eelet.org.uk

is crucial for developing targeted marketing strategies. For instance, a customer who frequently purchases and spends significantly but has not made a recent purchase may be targeted with re-engagement campaigns. Conversely, customers who excel in all three RFM metrics might be prioritized for reward schemes and upselling initiatives. RFM segmentation facilitates a nuanced understanding of customer behavior, enabling personalized customer relationship management.

The combination of RFM segmentation and K-means clustering, validated using techniques such as the Silhouette Score, provides a robust and mathematically sound method for accurately categorizing clients based on their purchasing patterns. This approach not only identifies meaningful clusters but also aligns with corporate strategies, supporting targeted marketing efforts and improved customer relationship management.

Temporal Analysis

Temporal analysis involves examining customer behavior over a designated six-month period. By applying temporal filters to the dataset and calculating revenue for each customer, this analysis provides valuable insights into short-term spending patterns. This approach enhances the broader RFM segmentation by focusing specifically on the most recent customer activity.

- **Revenue Calculation**: The total revenue for each customer over the six-month period was computed by multiplying the quantity of each product by its unit price and aggregating these values for each customer's transactions within the specified timeframe.
- **Monthly Consolidation**: By consolidating revenue data on a monthly basis, we identified expenditure patterns over the six months. This method allows for the detection of trends, such as peaks or troughs in spending, which may be related to seasonal variations, marketing campaigns, or other external factors.
- **Customer Categorization**: Customers were categorized based on their total spending during this period, enabling an evaluation of the financial impact of different consumer groups. The distribution of spending was analyzed using advanced statistical techniques, including percentile rankings and mean/median comparisons.

The analysis revealed significant insights, such as the proportion of total revenue generated by high-spending customers. Identifiable seasonal trends, such as increased expenditure during certain months, are crucial for inventory management and marketing strategies. Recognizing high-value clients within this timeframe allowed for targeted marketing efforts, such as offering exclusive promotions or loyalty programs to those who made substantial expenditures over the six months.

CLTV Predictive Modeling

To forecast Customer Lifetime Value (CLTV) clusters, the study employs a range of supervised machine learning models, including RandomForest, AdaBoost, ExtraTrees, LGBM, and XGBoost. These models are trained using features derived from both the RFM analysis and the six-month temporal analysis. K-fold cross-validation, with \(K = 5 \), is applied to evaluate the performance of each model. The accuracy results for all models are illustrated in Figure 1 below.

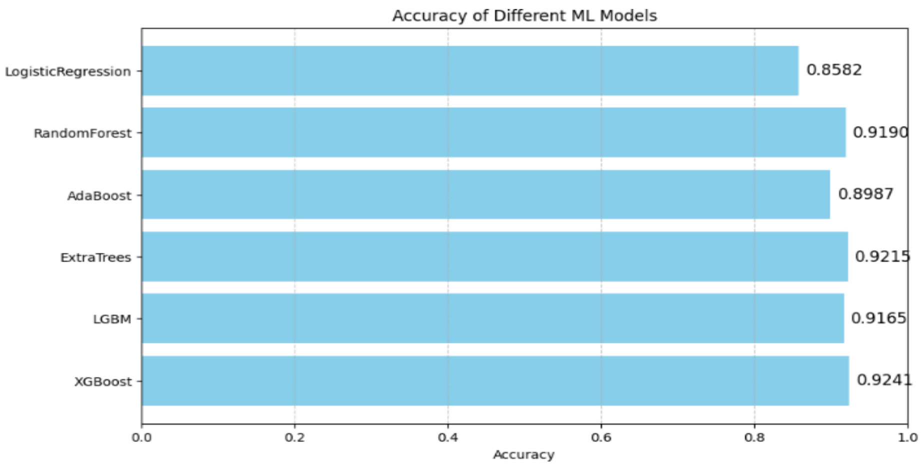


Fig. 1. ML Models Accuracy Comparison

The evaluation of model performance identifies the most effective models for forecasting Customer Lifetime Value (CLTV). For instance, models such as LGBM and XGBoost may demonstrate higher accuracy and AUC scores, reflecting their capability to distinguish between different consumer groups. Detailed results, including all evaluation metrics and machine learning models, are presented in Table 2 below.

Table 2. Evaluation Measures Results for ML Models

Model	Accuracy	Precision	Recall	F1_Score	AUC
LR	0.85822	0.80613	0.858228	0.812393	0.841505
RF	0.918987	0.913839	0.918987	0.914861	0.952288
AB	0.898734	0.87814	0.898734	0.888231	0.961658
ET	0.921519	0.914991	0.921519	0.915425	0.973859
LGBM	0.916456	0.913523	0.916456	0.914482	0.975394
XGB	0.924051	0.922711	0.924051	0.922893	0.973959

The insights obtained from this performance evaluation are critical for selecting the most suitable model for deployment in a business setting. A model with high precision and recall is particularly beneficial for marketing campaigns targeting high-value clients, as it ensures optimal resource allocation. Additionally, understanding model performance facilitates further refinement, such as adjusting parameters or exploring alternative feature sets, to enhance forecasting accuracy and improve business outcomes.

In our experiments, we employed a consistent training/testing ratio across all machine learning models to ensure a fair comparison. Specifically, an 80% training data and 20% testing data split was used, adhering to standard practices in predictive modeling. This uniform approach across RandomForest, AdaBoost, ExtraTrees, LGBM, and XGBoost ensures that our performance metrics are directly comparable and reliable.

Although our models demonstrate high accuracy in predicting Customer Lifetime Value (CLTV), we acknowledge the lack of direct benchmarking against state-of-the-art models on the same dataset. Future research should address this gap by comparing our methodology with existing approaches to better contextualize our findings within the broader landscape of retail analytics research.

5. Conclusion and Future Scope

The comprehensive analysis of customer behavior using a dataset from the Indian retail sector has achieved notable success in the domain of predictive analytics. By integrating RFM segmentation with temporal analysis and employing a rigorous data preprocessing and feature engineering pipeline, we have uncovered intricate patterns in customer purchasing behavior. The pinnacle of our analysis was reached with the application of advanced machine learning techniques, with the XGBClassifier emerging as the most effective model. This model demonstrated exceptional performance, achieving an accuracy of 92.40%, precision of 92.27%, recall of 92.40%, an F1 score of 92.28%, and an AUC score of 97.39%. The XGBClassifier not only validated the effectiveness of our methodology but also set a benchmark for predicting customer lifetime value clusters.

This research offers a thorough perspective on customer behavior by combining both descriptive and predictive analyses, providing valuable insights into the Indian retail sector. The findings have significant implications for enhancing customer relationship management and tailoring marketing strategies to maximize effectiveness and efficiency. The methodology and results establish a pioneering standard in the field, serving as a model for future studies that aim to utilize data analytics to understand and forecast customer behavior in the retail industry. Looking ahead, there is considerable potential to expand this research. Future studies should explore the temporal dynamics of customer behavior over extended periods, incorporating a broader range of variables, particularly those related to consumer demographics and multi-channel engagement metrics. Delving into deep learning could yield transformative insights, especially in processing and deriving value from unstructured data formats. Additionally, implementing dynamic segmentation strategies that adapt to recent customer interactions could revolutionize real-time consumer data analysis. Testing these approaches across different sectors will validate their adaptability and enhance their utility in various market environments. Future research should also assess the scalability and generalizability of these models to other regions, ensuring the continued efficacy of our analytics framework in diverse retail settings and expanding its scope and applicability.

References

1. S. H. Shihab, S. Afroge, and S. Z. Mishu, "RFM based market segmentation approach using advanced k-means and agglomerative clustering: a comparative study," in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1-4: IEEE.
2. M. Y. Smaili and H. J. A. S. E. J. Hachimi, "New RFM-D classification model for improving customer analysis and response prediction," p. 102254, 2023.
3. A. J. N. C. Alsayat and Applications, "Customer decision-making analysis based on big social data using machine learning: a case study of hotels in Mecca," vol. 35, no. 6, pp. 4701-4722, 2023.
4. D. Bratina and A. J. M.-T. Faganel, "Using Supervised Machine Learning Methods for RFM Segmentation: A Casino Direct Marketing Communication Case," vol. 35, no. 1, pp. 7-22, 2023.
5. T. F. Bahari and M. S. J. P. c. s. Elayidom, "An efficient CRM-data mining framework for the prediction of customer behaviour," vol. 46, pp. 725-731, 2015.
6. I. Lewaa, M. S. Hafez, and M. A. J. S. J. o. t. I. Ismail, "Data integration using statistical matching techniques: A review," vol. 37, no. 4, pp. 1391-1410, 2021.
7. I. Lewaa, M. S. Hafez, and M. A. Ismail, "Mixed Statistical Matching Approaches Using a Latent Class Model: Simulation Studies," 2023.
8. M. A. Rahim, M. Mushafiq, S. Khan, Z. A. J. J. o. R. Arain, and C. Services, "RFM-based repurchase behavior for customer classification and segmentation," vol. 61, p. 102566, 2021.
9. J. Mohammad and M. A. Kashem, "Air Pollution Comparison RFM Model Using Machine Learning Approach," in 2022 IEEE 7th International conference for Convergence in Technology (I2CT), 2022, pp. 1-5: IEEE.
10. T. Jiang, A. J. I. t. o. k. Tuzhilin, and d. engineering, "Improving personalization solutions through optimal segmentation of customer bases," vol. 21, no. 3, pp. 305-320, 2008.
11. S. Shah and M. Singh, "Comparison of a time efficient modified K-mean algorithm with K-mean and K-medoid algorithm," in 2012 international conference on communication systems and network technologies, 2012, pp. 435-437: IEEE.
12. Y. S. Cho and S. C. J. J. Moon, "Weighted mining frequent pattern based customer's RFM score for personalized u-commerce recommendation system," vol. 4, no. 4, pp. 36-40, 2013.
13. N. Lu, H. Lin, J. Lu, and G. J. I. T. o. I. I. Zhang, "A customer churn prediction model in telecom industry using boosting," vol. 10, no. 2, pp. 1659-1665, 2012.
14. X. He and C. Li, "The research and application of customer segmentation on e-commerce websites," in 2016 6th International Conference on Digital Home (ICDH), 2016, pp. 203-208: IEEE.
15. A. Sheshasaayee and L. Logeshwari, "An efficiency analysis on the TPA clustering methods for intelligent customer segmentation," in 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2017, pp. 784-788: IEEE.
16. R. Parvez, Tarantino, A., and Meerza, S. I. A. Understanding the Prediction of Student Retention Behavior during COVID-19 Using Effective Data Mining Techniques. In Research Square - Preprint, 2023. DOI: 10.21203/rs.3.rs-2948727/v1.
17. S. Arefin, R. Parvez, T. Ahmed, F. Jahin, F. Sumaiya, and M. Ahsan, "Enhancing Bank Consumer Retention: A Comprehensive Analysis of Churn Prediction Using Machine Learning and Deep Learning Techniques," in Proceedings of the 5th International Conference on Advances in Information Communication Technology & Computing (AICTC-2024), April 29-30, 2024. doi: 10.13140/RG.2.2.18210.85443.
18. R. Parvez, Meerza, S. I. A., and Khan Chowdhury, N.H. Economics of student retention behavior in higher education. In Proceedings of the Annual Meeting—Agricultural and Applied Economics Association (AAEA), Kansas City, Missouri, 26–28 July 2020.

19. S. Arefin, R. Parvez, M. Rahman, F. Sumaiya, N. H. Khan Chowdhury, and M. Ahsan, “Understanding Bankruptcy Prediction Using Data Mining Algorithms - Evidence from Taiwan's Economy,” in Proceedings of the 9th International Congress ICICT, London, United Kingdom, 19-22 February 2024. doi: 10.13140/RG.2.2.17919.21925.
20. R. Parvez, Meerza, S. I. A., and Khan Chowdhury, N.H. Forecasting student enrollment using time series models and recurrent neural networks. In Proceedings of the Joint Annual Meeting— Agricultural and Applied Economics Association (AAEA) & Western Agricultural Economics Association (WAEA), Austin, TX 1–3 August 2021.
21. R. Parvez, N. H. Khan Chowdhury, A.J. Smart, L.L. Janssen, and R.G. Stover, “Conservation Attitudes and Satisfaction on Agro-Environmental Conditions of South Dakota: A study of Landowners’ Perceptions,” Journal of the American Society of Farm Managers and Rural Appraisers, vol. 3, no. 1, pp. 11-25, 2017.
22. M. R. Parvez and L.L. Janssen, "Assessing Producer Perceptions and Activities on Agro- Environment Conditions: Sociol-Economic and Behavioral Dimensions," in 2011 48th Missouri Valley Economic Association Annual Meeting, 2011, MVEA.
23. Ahsan, M., Gomes, R., Chowdhury, M.M. and Nygard, K.E., 2021. Enhancing machine learning prediction in cybersecurity using dynamic feature selector. Journal of Cybersecurity and Privacy, 1(1), pp.199-218.
24. Ahsan, M., Gomes, R. and Denton, A., 2018, May. Smote implementation on phishing data to enhance cybersecurity. In 2018 IEEE international conference on electro/information technology (EIT) (pp. 0531-0536). IEEE.
25. Rifat, N., Ahsan, M., Chowdhury, M. and Gomes, R., 2022, May. BERT against social engineering attack: Phishing text detection. In 2022 IEEE International Conference on Electro Information Technology (eIT) (pp. 1-6). IEEE.
26. Ahsan, M.K., 2021. Increasing the predictive potential of machine learning models for enhancing cybersecurity (Doctoral dissertation, North Dakota State University).
27. Rifat, N., Ahsan, M., Gomes, R. and Chowdhury, M., 2022, May. Covid-19 sentiment analysis applying bert. In 2022 IEEE International Conference on Electro Information Technology (eIT) (pp. 417-422). IEEE.
28. Chowdhury, M., Rifat, N., Latif, S., Ahsan, M., Rahman, M.S. and Gomes, R., 2023, May. ChatGPT: The Curious Case of Attack Vectors' Supply Chain Management Improvement. In 2023 IEEE International Conference on Electro Information Technology (eIT) (pp. 499-504). IEEE.
29. Gomes, R., Ahsan, M. and Denton, A., 2018, May. Random forest classifier in SDN framework for user-based indoor localization. In 2018 IEEE International Conference on Electro/Information Technology (EIT) (pp. 0537-0542). IEEE.
30. S. Arefin, R. Parvez, T. Ahmed, F. Jahin, F. Sumaiya, and M. Ahsan, “Enhancing Bank Consumer Retention: A Comprehensive Analysis of Churn Prediction Using Machine Learning and Deep Learning Techniques,” in Proceedings of the 5th International Conference on Advances in Information Communication Technology & Computing (AICTC-2024), April 29-30, 2024. doi: 10.13140/RG.2.2.18210.85443.