

Comparison of Classification Algorithm Performance in Predicting Customer Churn in Retail Stores Based on the RFM Model



Mohammad Ghasemi Zadeh¹, Seyed Abdollah Amin Mousavi^{2,*} and Mohammad Maleki Nia³

¹ Department of Information Technology Management, KI.C., Islamic Azad University, Kish, Iran; 

² Department of Management, CT.C., Islamic Azad University, Tehran, Iran; 

³ Department of Information Technology Management, ST.C., Islamic Azad University, Tehran, Iran; 

* Correspondence: saa.mousavi@iau.ac.ir

Citation: Ghasemi Zadeh, M., Mousavi, S. A. A., & Maleki Nia, M. (2027). Comparison of Classification Algorithm Performance in Predicting Customer Churn in Retail Stores Based on the RFM Model, *Business, Marketing, and Finance Open*, 4(1), 1-18.

Received: 14 December 2025

Revised: 01 May 2026

Accepted: 08 May 2026

Initial Publication: 07 October 2026

Final Publication: 01 January 2027



Copyright: © 2027 by the authors. Published under the terms and conditions of Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Abstract: Predicting customer churn is considered one of the fundamental issues in the retail industry because a decline in customer retention rates directly affects acquisition costs and profitability. Accordingly, the present study was conducted with the aim of developing a comprehensive analytical framework for customer churn prediction based on the behavioral RFM model (Recency, Frequency, Monetary) by utilizing various machine learning algorithms. In this study, actual customer transaction data from a retail store were used, and the algorithms of Logistic Regression, Decision Tree, Support Vector Machine (SVM), Naïve Bayes, AdaBoost, and the hybrid Stacking model were evaluated and compared. Following normalization, data cleaning, and feature engineering processes, the models were assessed using the quantitative metrics of Precision, Accuracy, Recall, and F1-Score. The evaluation results demonstrated that the Stacking model outperformed the other algorithms and, with an overall accuracy of approximately 87%, provided a highly accurate framework for predicting churn probability. Variable importance analysis indicated that the Recency indicator plays the most significant role in predicting churn tendency, while the combination of Frequency and Monetary indicators improves model stability. In addition, K-Means clustering and association rule techniques (FP-Growth) were employed for behavioral analysis and customer segmentation, and the results emphasized the effectiveness of combining purchasing pattern analysis with RFM segmentation to improve churn prediction model accuracy. The primary innovation of this study lies in the integration of three analytical approaches—association rules, Stacking classification, and RFM clustering—within a unified empirical framework that establishes a systematic relationship among customer segmentation, purchasing behavior, and churn probability. This framework can serve as a practical basis for designing intelligent customer retention campaigns in the retail industry.

Keywords: Customer churn prediction, RFM model, Stacking algorithm, retail data mining, association rules (FP-Growth), K-Means clustering, machine learning, customer loyalty

1. Introduction

Customer churn has become one of the most critical analytical concerns in contemporary retail management because customer retention, repeat purchasing, and long-term profitability are increasingly shaped by the ability of firms to identify customers who are likely to reduce or terminate their relationship with the business. In highly competitive retail environments, the cost of acquiring new customers is typically higher than the cost of retaining

existing ones, and therefore churn prediction is not merely a technical modeling task but a strategic decision-support mechanism for customer relationship management, marketing resource allocation, and loyalty program design. The expansion of digital retail channels, omni-channel purchasing behavior, and data-rich transaction systems has intensified the importance of predictive analytics in understanding how customers move across channels, respond to marketing stimuli, and demonstrate changing levels of loyalty over time [1]. In this context, customer churn prediction enables retailers to move from reactive decision-making toward proactive customer retention strategies by detecting behavioral signals before customer disengagement becomes irreversible.

The retail sector is especially suitable for churn analytics because customers continuously generate transactional, temporal, and behavioral data through purchases, basket composition, purchase frequency, spending value, and intervals between visits. Unlike contractual industries such as telecommunications or banking, where churn may be explicitly observable through account termination, retail churn is often implicit and must be inferred from changes in purchasing behavior. This makes retail churn prediction methodologically more complex, because the absence of a purchase does not always indicate permanent defection; rather, it may reflect delayed repurchase, seasonal demand variation, channel switching, or changes in household consumption patterns. Earlier studies on churn prediction have emphasized that machine learning techniques can identify complex patterns in customer behavior that are difficult to detect through traditional statistical approaches [2]. Accordingly, retail churn modeling requires a structured representation of customer behavior and robust classification algorithms capable of distinguishing between active, inactive, valuable, and at-risk customers.

Among the behavioral models used in customer analytics, the Recency, Frequency, and Monetary model, commonly known as RFM, has remained one of the most practical and interpretable frameworks for customer segmentation and churn analysis. Recency reflects the time elapsed since the customer's most recent purchase, Frequency represents the number of purchases made during a defined period, and Monetary value measures the total financial contribution of the customer. These indicators provide a compact but powerful summary of purchasing behavior and are widely used for customer classification, segmentation, and repurchase prediction. Research on RFM-based customer classification has shown that these three dimensions can effectively differentiate customers according to their purchasing potential, loyalty level, and probability of future engagement [3]. In retail contexts, RFM indicators are particularly valuable because they can be extracted directly from transaction records and converted into input variables for machine learning models without requiring extensive attitudinal or demographic data.

The integration of RFM analysis with machine learning has opened new opportunities for improving the accuracy and practical relevance of customer churn prediction. Traditional RFM applications often rely on scoring rules or heuristic segmentation, whereas machine learning models allow the RFM structure to be used as a predictive feature space for classification, clustering, and behavioral modeling. Recent studies have demonstrated that machine learning-based RFM analysis can support more precise customer segmentation by identifying hidden patterns in customer value and purchase behavior [4]. Similarly, comparative work on cluster-based churn prediction has shown that combining RFM metrics with data mining techniques can enhance the identification of meaningful customer groups compared with conventional RFM-based models alone [5]. These findings indicate that RFM should not be viewed only as a descriptive tool but as a foundational behavioral representation that can be integrated into more advanced predictive systems.

Machine learning has been extensively applied to churn prediction because of its ability to process large-scale customer datasets and model nonlinear relationships among behavioral variables. Supervised classification

algorithms such as logistic regression, decision trees, support vector machines, k-nearest neighbors, naïve Bayes, neural networks, and ensemble methods have been used to predict whether customers are likely to churn or remain active. Comparative studies of supervised machine learning techniques have shown that model performance varies depending on data structure, feature selection, preprocessing strategy, class imbalance, and evaluation criteria [6]. Therefore, no single algorithm can be assumed to be universally superior across all customer churn settings. Instead, algorithmic performance must be evaluated empirically under comparable data conditions using standardized metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve.

In recent years, ensemble learning methods have gained increasing attention in churn prediction research because they combine multiple learners to improve predictive stability and reduce model-specific weaknesses. Methods such as bagging, boosting, stacking, and integrated learning are designed to enhance classification performance by aggregating diverse decision boundaries or sequentially correcting prediction errors. Research on online shopping churn prediction has highlighted the potential of integrated learning approaches for improving predictive accuracy in digital retail environments [7]. More recent comparative analyses have also emphasized that ensemble methods represent a strong state-of-the-art approach in customer churn prediction because they often outperform individual base learners when appropriately configured [8]. In particular, stacking classifiers have attracted attention because they combine heterogeneous models and use a meta-learner to optimize final predictions, making them suitable for complex behavioral datasets where different algorithms capture different aspects of churn behavior [9].

Despite the promise of ensemble models, classical algorithms remain important in churn research because of their interpretability, computational efficiency, and managerial usability. Logistic regression provides clear coefficient-based explanations of churn probability, decision trees offer rule-based interpretability, naïve Bayes performs effectively under certain probabilistic assumptions, and support vector machines can handle high-dimensional boundaries. Customer churn studies have shown that predictive models should not be assessed only on accuracy, because managerial decision-making also depends on the ability to correctly identify churners and minimize false positives or false negatives [10]. For example, a model with high recall may be valuable when the firm prioritizes identifying as many at-risk customers as possible, while a model with high precision may be preferable when retention campaigns are costly and should be targeted only at customers with a high probability of churn. Thus, comparing multiple algorithms across several performance metrics is essential for selecting a model that is both statistically effective and managerially useful.

Deep learning has also entered customer churn and retail behavior prediction as a response to increasingly complex customer journeys and sequential purchase patterns. Deep neural networks are capable of modeling nonlinear interactions among variables and can capture hidden representations in large behavioral datasets. Studies on retail consumer behavior have demonstrated the potential of deep neural networks for predicting next logical purchases and modeling consumer decision patterns in data-rich environments [11]. Similarly, sequential deep learning frameworks have been proposed for retail churn modeling, particularly when the objective is to capture the order, timing, and evolution of customer transactions over time [12]. In the fast-moving consumer goods segment, deep learning has also been applied to churn prediction because FMCG customers often exhibit frequent, repetitive, and category-dependent purchase behavior that may require advanced pattern recognition [13]. However, deep learning models may require larger datasets, greater computational resources, and more careful interpretability mechanisms than classical or ensemble classifiers.

The application of churn prediction is not limited to retail; evidence from telecommunications, banking, e-commerce, and FMCG sectors has contributed to the development of churn modeling methods that can be adapted to retail settings. For instance, studies in the telecommunications sector have shown that machine learning techniques can effectively identify customers at risk of leaving when behavioral and usage-related variables are available [14]. In retail banking, churn prediction frameworks have emphasized the need to improve predictive performance through careful feature engineering, model selection, and validation procedures [15]. Although these sectors differ from retail stores in customer relationship structure, they provide important methodological lessons regarding classification design, evaluation metrics, and predictive model comparison. Retail churn prediction can benefit from these insights while adapting the definition of churn to the specific characteristics of non-contractual purchasing behavior.

Another important direction in customer analytics is the combination of predictive and descriptive data mining techniques. Classification models predict whether customers will churn, while clustering models identify homogeneous customer segments and association rule mining reveals relationships among purchased items. A customer behavior mining framework can integrate clustering and classification techniques to provide a more comprehensive understanding of customer behavior, moving beyond single-model prediction toward broader decision support [16]. In retail management, this integration is highly relevant because churn is not only a binary outcome but also a behavioral process connected to customer value, product preference, basket composition, and purchase regularity. Accordingly, combining RFM-based segmentation with classification and association rule analysis can help retailers understand not only which customers are likely to churn but also what products they purchase, which customer groups they belong to, and which marketing actions may retain them.

From a managerial perspective, churn prediction is meaningful only when it can inform actionable retention strategies. Customer lifetime value, retention, and churn are interconnected constructs that determine how firms should allocate marketing resources across customer segments [17]. High-value customers who show early signs of disengagement may require personalized incentives, loyalty reinforcement, or product recommendations, whereas low-value or inactive customers may require different engagement strategies. In retail stores, this requires linking churn prediction outputs with customer segmentation, market basket analysis, and targeted campaign planning. The identification of valuable non-churning customers, at-risk customers, and product co-purchase patterns can support more precise promotional design, improved store layout, cross-selling strategies, and customer loyalty programs. Therefore, churn prediction should be positioned as part of an integrated retail intelligence system rather than a standalone classification exercise.

The methodological challenge in this field lies in designing a predictive framework that is both empirically rigorous and practically interpretable. A robust churn prediction study must include data preprocessing, feature engineering, normalization, model training, model comparison, cross-validation, and evaluation based on multiple performance indicators. It must also address the specific nature of retail data, including high transaction volume, repeated purchases, heterogeneous product categories, irregular purchasing intervals, and implicit churn labels. Recent work on customer churn prediction based on machine learning algorithms has emphasized the importance of testing multiple classifiers to determine which algorithm best fits the given data structure [18]. At the same time, the retail environment requires models that can be translated into operational decisions, especially when the goal is to increase customer retention, improve customer lifetime value, and design evidence-based marketing interventions.

Based on these considerations, the present study focuses on predicting customer churn in a retail store using RFM-based behavioral indicators and comparing the performance of multiple machine learning classification algorithms. The study is situated at the intersection of customer behavior analytics, retail data mining, and predictive modeling. It uses actual retail transaction data, derives RFM indicators, applies preprocessing and normalization procedures, and evaluates several classical and ensemble classification algorithms under comparable conditions. In addition to predictive classification, the broader analytical framework also considers the relevance of clustering and association rule mining for interpreting customer behavior and identifying product relationships. This approach responds to the growing need for retail decision-support systems that combine predictive accuracy with managerial interpretability.

Therefore, the aim of this study is to compare the performance of selected machine learning classification algorithms in predicting customer churn in a retail store based on RFM indicators and to identify the most effective model for supporting customer retention and data-driven retail decision-making.

2. Methodology

This study was designed to compare the performance of a set of classification algorithms in predicting customer churn in retail stores based on the RFM model. Instead of using a hybrid approach combining association rules and clustering techniques, the study focused entirely on the classification process in order to evaluate the generalizability and accuracy of the algorithms under identical conditions. The research framework was implemented according to the standard CRISP-DM process, including data preparation, data understanding, modeling, and evaluation stages, with the distinction that the workflow was specifically reconstructed for churn evaluation using RFM indicators. In summary, raw data were first prepared and cleaned, then entered into the modeling stage, and finally the performance of each algorithm was assessed using evaluation metrics.

The data used in this study were derived from actual customer transactions of a retail store over a specified time period. The dataset included the date of the latest purchase, customer ID, total purchase amount, number of purchases, and time-related attributes such as month, season, and weekday. The first step involved calculating the RFM indicators, including Recency (number of days since the last purchase), Frequency (number of purchases within the specified period), and Monetary value (total amount paid by the customer during the period).

To prepare the data, the following stages were continuously performed: removing duplicate and incomplete records, feature engineering to extract RFM indicators and encode time variables, normalizing indicator ranges using the Min-Max technique, and assigning churn/non-churn labels based on a predefined inactivity threshold within a specific period. Subsequently, the data were divided into training (70%) and testing (30%) sets to ensure out-of-sample evaluation. A description of the raw data structure and a sample of customer transaction records are presented in Table 1, including all input features used in the classification models.

Table 1. A Sample of Customer Transaction Datasets from a Retail Store

Store Name	Product Name	Product Barcode	Customer ID	Season	Month	Day of Week	Day	Net Sales Amount After Returns (IRR)	Last Sales Date	Days Since Last Purchase	Total Quantity Sold	Number of Sales Invoices
Seraj Branch 1	Saei Low-Absorption Frying Oil 810 g	6260295101705	9010222018	Summer	August	Thursday	20	613,040	2022-08-11	–	1	1
Seraj Branch 1	Softlan Golden Automatic Laundry	6260010500394	9010222018	Summer	August	Thursday	20	153,546	2022-08-11	–	1	1

	Powder 500 g (24-pack)											
Seraj Branch 1	Pegah Domestic Butter 25 g	6260707400587	9010222018	Summer	August	Thursday	20	59,037	2022- 08-11	–	1	1
Seraj Branch 1	Schon Fruity Mix Shampoo 400 ml	6260482526373	9010222018	Summer	August	Thursday	20	436,055	2022- 08-11	–	1	1
Seraj Branch 1	Camelion AA Batteries (2-pack)	4260033151032	9010222018	Summer	August	Thursday	20	90,826	2022- 08-11	–	1	1
Seraj Branch 1	Signal Anti- Cavity Toothpaste 100 ml	6260526418176	9010222018	Summer	August	Thursday	20	234,862	2022- 08-11	–	1	1
Seraj Branch 1	Kaleh White UF Cheese 400 g	6260161529220	9010222018	Summer	August	Thursday	20	288,000	2022- 08-11	–	1	1

To implement the modeling phase, a collection of ensemble and classical classification algorithms was selected, including Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree, AdaBoost, and Stacking. Each algorithm, considering the research objectives and data structure, provides a distinct approach to identifying customer churn patterns. The algorithms were selected based on model diversity, interpretability, and effectiveness in previous related studies.

Implementation was carried out in the Python environment using libraries such as Pandas and NumPy for data management, Scikit-learn for model implementation, and Seaborn and Matplotlib for result visualization. In the results section, the algorithms are ranked based on evaluation metrics to determine the best-performing model.

In this study, four primary metrics were considered to evaluate model performance: Accuracy, Precision, Recall, and F1-Score. Each metric evaluates a different aspect of classifier performance. For example, Accuracy indicates the ratio of correct predictions to total predictions, whereas Recall measures the model's capability to identify all churned customers. To improve result reliability, in addition to evaluation on the testing dataset, 10-fold cross-validation was applied. The overall workflow of the study, from data collection to final evaluation, is illustrated in Figure 1, where the flow of data from source to final model outputs is depicted. This diagram demonstrates the interdependence of results and the continuity of each stage with data quality and proper model selection.

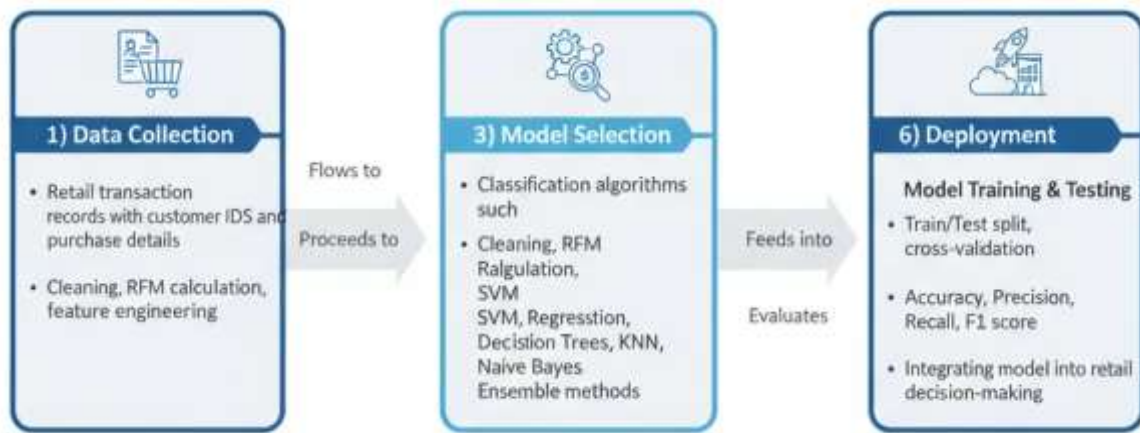


Figure 1. Flowchart of the Customer Churn Classification Process Based on RFM

In this section, the implementation pathway of the study, from framework development to selecting the optimal classification algorithm, was explained. Transactional data were enriched using RFM indicators and, after preprocessing, entered into the modeling stage using selected algorithms. To ensure comprehensive analysis, a set

of standard metrics and indicators was employed, and the research workflow was presented in the form of a process flowchart.

3. Findings and Results

Analyzing and evaluating the performance of machine learning models in customer churn prediction requires a combination of quantitative assessment and qualitative interpretation of results. The purpose of this section is to explain the applied approaches, identify the strengths and weaknesses of each model, and provide a clear representation of the data processing and model development procedures. The investigated models include ensemble algorithms such as Bagging, Stacking, and AdaBoost, classical algorithms such as Decision Tree (DT), Logistic Regression (LR), KNN, Support Vector Machine (SVM), Naïve Bayes (NB), and Artificial Neural Networks. Each model was evaluated based on Precision, Accuracy, Recall, and F1-Score to compare their abilities in identifying churned and non-churned customers. The actual data used in this study were extracted from the sales system of a retail store, and after preprocessing, the RFM model indicators were calculated.

The data used in this study were extracted from actual customer transactions of Seraj Store Branch 1 over a specified period from the store sales system. The dataset contained approximately 541,000 records and 13 features related to the purchasing process, where each record represented an independent transaction (Table 1). The primary objective of this section is to analyze and evaluate the distribution of key customer purchasing behavior indicators based on the RFM model. After preprocessing and removing unnecessary features (such as store name, season, and month of purchase), the data were transformed into a standardized format suitable for association rule mining, clustering, and classification models. The distribution of these three indicators reveals important customer purchasing behavior patterns, which are visually illustrated below. As shown in Table 1, a sample of customer transaction data includes features such as product barcode, product name, purchase amount, and customer ID. Based on these data, the RFM distribution was plotted as shown in Figure 2.



Figure 2. Distribution of RFM Metrics Based on Actual Data from Seraj Store Branch 1

As illustrated in Figure 2, the chart includes three histograms representing the distributions of Recency, Frequency, and Monetary, respectively. A substantial proportion of customers exhibit low Recency values (recent purchases), whereas the Monetary and Frequency distributions are heterogeneous and multimodal, indicating the presence of different purchasing behavior patterns.

In this section, the performance of various classification models used in this study for customer churn prediction is examined. The data described in the previous section, after preprocessing and transformation, were input into various machine learning algorithms, including ensemble and classical techniques. These models were evaluated based on four primary criteria: Accuracy, Precision, Recall, and F1-Score. According to the results presented in

Tables 2 and 3, the Stacking algorithm achieved the best overall performance ranking, followed by Logistic Regression and Naïve Bayes in second and third positions, respectively.

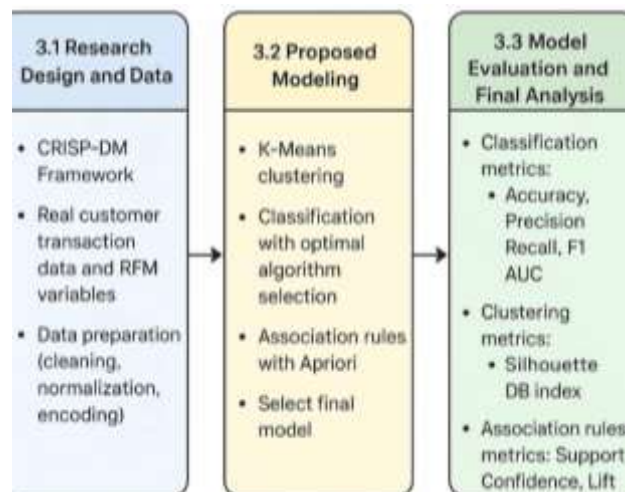
Table 2. Ranking of Classification Algorithms for Predicting Customer Churn in Retail

Classification Algorithm	Accuracy	Precision	Recall	F1-Score	Rank
Stacking	86.47	84.07	90.00	86.93	1
Logistic Regression	85.88	79.90	95.88	87.17	2
Naïve Bayes	85.59	83.06	89.41	86.12	3
Support Vector Machine	84.71	82.07	88.82	85.31	4
Decision Tree	83.24	83.43	82.94	83.19	5
AdaBoost	84.12	81.87	87.65	84.66	6
Bagging	72.35	64.62	98.82	78.14	7
Neural Network	81.18	77.32	88.24	82.42	8
K-Nearest Neighbors	66.76	61.83	87.65	72.51	9
Linear Discriminant Analysis	58.53	57.92	62.35	60.06	10

Table 3. Comparison of Classification Algorithms Based on Index Ranking

Classification Algorithm	Accuracy	Precision	Recall	F1-Score	Average Rank
Decision Tree	6	2	9	6	5.75
K-Nearest Neighbors	9	9	7	9	8.50
Naïve Bayes	3	3	4	3	3.25
Neural Network	7	7	6	7	6.75
Logistic Regression	2	6	2	1	2.75
Support Vector Machine	4	4	5	4	4.25
Linear Discriminant Analysis	10	10	10	10	10.00
AdaBoost	5	5	8	5	5.75
Bagging	8	8	1	8	6.25
Stacking	1	1	3	2	1.75

Following the quantitative evaluation of classification model performance metrics, the results were also visually represented through graphical charts to facilitate comparison and interpretation. These charts clearly illustrate the differences in predictive capability and classification accuracy among the algorithms. Figure 3 presents the evaluation framework of the models based on key indicators, while Figure 4 displays the ROC curves for comparing the efficiency of selected algorithms in customer churn prediction.

**Figure 3. Overall Framework for Evaluating Data Mining Models Based on RFM Metrics**

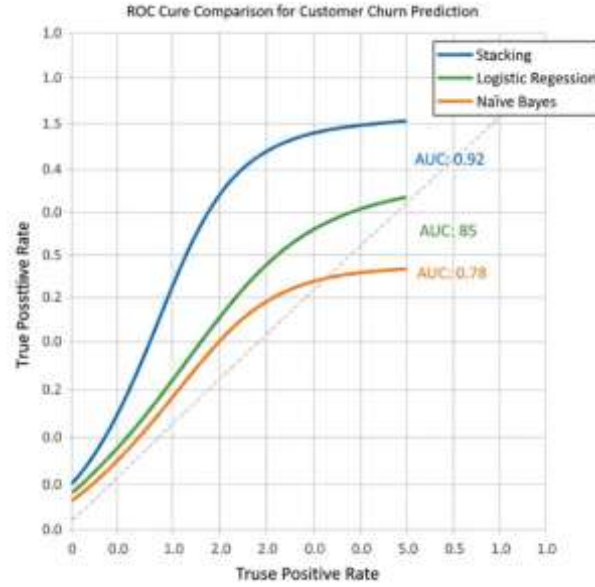


Figure 4. Comparison of ROC Curves in Predicting Customer Churn for a Retail Store

Figure 3 illustrates the three major stages of the analytical process. In the first stage, actual customer data and RFM indicators are normalized and prepared according to the CRISP-DM framework. In the second stage, modeling is performed using three complementary perspectives: K-Means clustering, Apriori rule extraction, and classification with optimal algorithm selection. Finally, the models are evaluated using a collection of metrics, including Accuracy, Recall, Precision, F1, and AUC for classification, DB Index and Silhouette for clustering, and Support, Lift, and Confidence for association rules. This framework visually demonstrates the complete data mining process from data preparation to final analysis and interpretation.

Figure 4 presents the ROC performance curves of the three algorithms: Stacking, Naïve Bayes, and Logistic Regression in correctly identifying churned customers. The horizontal axis represents the False Positive Rate, while the vertical axis represents the True Positive Rate. Furthermore, the Area Under the Curve (AUC) serves as an indicator of the discriminative power of the models. According to the results, the Stacking model achieved the highest AUC (approximately 0.92), demonstrating superior capability in distinguishing churned from non-churned customers compared to the other two models. In contrast, the Naïve Bayes model, with an AUC of approximately 0.78, exhibited weaker performance. This comparison highlights the superiority of the ensemble-based Stacking approach in customer churn analysis and evaluation.

In this section, two descriptive data mining techniques, namely K-Means clustering and Association Rules, were employed to uncover hidden relationships among purchased items and identify customer behavioral patterns. This analysis was conducted using actual transaction data from Seraj Store Branch 1.

In the first step, the FP-Growth algorithm was implemented in the RapidMiner environment to identify products with the highest co-purchase frequency. According to the process illustrated in Figure 5, the data were first preprocessed and textual features were transformed into categorical data to prepare the dataset for extracting strong association rules. In this algorithm, the minimum thresholds for Confidence and Support were both set at 0.5 to ensure that only the most frequent purchasing patterns were extracted.



Figure 5. Application of Association Rules for Market Basket Analysis

The results obtained from model implementation are presented through the outputs shown in Tables 4 to 7. The analysis and evaluation of these results indicate that certain specific items repeatedly appear together in the shopping baskets of non-churning customers. For example, combinations of items with the codes 2200170, 2203314, 2202312, and 2204359 have the highest support values. The gradual decrease in support values from four-item combinations to three-item and then two-item combinations indicates that the probability of the simultaneous occurrence of larger itemsets is lower.

Table 4. Top 4 Frequently Purchased Items in Transactions of Valuable Non-Churning Customers

Size	Support	Item 1	Item 2	Item 3	Item 4
4	0.121	2200170	2203314	2202312	2204359
4	0.127	2200170	2203314	2202312	2204364

Table 5. Top 3 Frequently Purchased Items in Transactions of Valuable Non-Churning Customers

Size	Support	Item 1	Item 2	Item 3
3	0.139	2200170	2203314	2202312
3	0.121	2200170	2203314	2204359
3	0.128	2200170	2203314	2204364
3	0.121	2200170	2202312	2204359
3	0.127	2200170	2202312	2204364
3	0.122	2203314	2202312	2204359
3	0.128	2203314	2202312	2204364

Table 6. Top 2 Frequently Purchased Items in Transactions of Valuable Non-Churning Customers

Size	Support	Item 1	Item 2
2	0.117	2201035	6260055843036
2	0.140	2200170	2203314
2	0.139	2200170	2202312
2	0.123	2200170	2204359
2	0.128	2200170	2204364
2	0.141	2203314	2202312
2	0.125	2203314	2204359
2	0.130	2203314	2204364
2	0.124	2202312	2204359
2	0.128	2202312	2204364

Table 7. Top 1 Frequently Purchased Item in Transactions of Valuable Non-Churning Customers

Size	Support	Item 1
1	0.332	2201035
1	0.295	6260055843036
1	0.220	6261906101176
1	0.194	2208121
1	0.192	2206861
1	0.190	6260010500394
1	0.186	6267604414693
1	0.176	164200275300044
1	0.176	2201946
1	0.174	164200275300001
1	0.171	6260043900017
1	0.160	6262916600741
1	0.156	2200134
1	0.153	6260016342011
1	0.152	2200170
1	0.152	2200917
1	0.151	2203314
1	0.149	6265499500040
1	0.149	164200275300002
1	0.149	2201157
1	0.147	6269523900059
1	0.147	2202312
1	0.146	6260459800079
1	0.142	6261847502209
1	0.142	6261061100533
1	0.141	6260053186241
1	0.141	2204359
1	0.141	6260176802929
1	0.138	2201000
1	0.138	6263812801249
1	0.138	164200203250232
1	0.136	2204364
1	0.131	6260105005155
1	0.129	2201655
1	0.129	2200893
1	0.126	6261985200050
1	0.126	6260161529275
1	0.125	6260532822097
1	0.125	2201691
1	0.120	6262157900105
1	0.120	6260667302532
1	0.120	6260053183974
1	0.119	6260707400587
1	0.119	2201014
1	0.119	6263812801386
1	0.117	6260667302525
1	0.117	6260053183998
1	0.117	2201015
1	0.116	6268187102113
1	0.115	6265547620287
1	0.115	2201016
1	0.115	2201658

To visualize the discovered relationships among frequent items, a network structure of association rules was plotted, showing the relationships between items based on the Lift index (Figure 6). In this network, nodes represent frequent products, and connecting lines indicate the strength of association among them. The greater the thickness of the edges, the stronger the dependency or co-purchase relationship among items. This diagram clearly identifies the key items that appear in a larger number of rules and demonstrates their role in customers' shopping baskets.

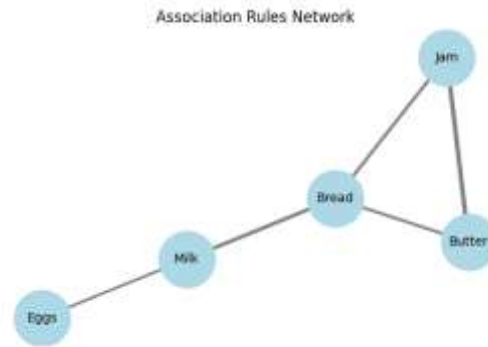


Figure 6. Association Network of Frequent Items Based on Association Rules

The patterns extracted from the FP-Growth algorithm indicate that some products have strong complementary relationships. From a marketing perspective, these relationships can be used to design combined sales bundles, product recommendation offers, or more effective product placement in the store. These results provide a reliable basis for targeted promotional planning and strengthening customer loyalty.

Subsequently, to complete the descriptive analysis, the K-Means algorithm was applied to segment customers into several homogeneous groups based on RFM indicators. The results of the algorithm showed that customers can be divided into three main clusters: high-value loyal customers, potential customers with moderate activity, and customers at risk of churn with lower activity. Consequently, each cluster has its own specific statistical characteristics, determined by the mean values of Recency, Frequency, and Monetary. As shown in Figure 7, the position of each point represents a specific customer; the vertical axis indicates the customer's Monetary value, and the horizontal axis indicates the customer's Recency, while the color of the points represents the cluster to which each customer belongs. As observed in Figure 7, loyal customers are concentrated in the upper region and closer to the lower Recency axis, meaning more recent purchases, and they have higher Monetary values. In contrast, customers at risk of churn are located in the lower section of the chart and are characterized by longer purchase intervals.

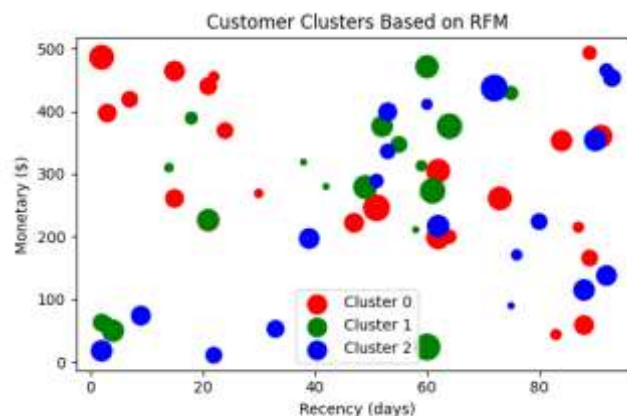


Figure 7. Scatter Plot of Customer Clusters Based on RFM Metrics

The examination of clusters showed that the frequent components of association rules mainly correspond to the potential and loyal customer clusters. This finding indicates the logical overlap of frequent purchasing patterns with more profitable market segments. The simultaneous analysis and evaluation of association rule and clustering results provide a comprehensive picture of customers' purchasing behavior and can serve as a basis for offering personalized recommendations and designing more effective loyalty campaigns. Overall, combining K-Means and FP-Growth techniques in this analysis made it possible to simultaneously identify customer behavioral characteristics and product relationships. The association rule algorithm revealed complementary relationships among products, while RFM clustering identified valuable customer segments. These results can be used in developing customer retention strategies, designing combined sales bundles, and planning intelligent promotional programs to increase the customer lifetime value of Seraj Store.

4. Discussion and Conclusion

The findings of the present study demonstrated that machine learning algorithms can effectively predict customer churn in retail environments when customer behavioral indicators are represented through the RFM framework. The results showed that the Stacking classifier achieved the highest overall performance among all evaluated models, followed by Logistic Regression and Naïve Bayes. The superiority of the Stacking approach indicates that ensemble-based learning methods provide stronger predictive capability in retail churn prediction because they combine the strengths of multiple classifiers and reduce the weaknesses associated with individual algorithms. This finding is consistent with previous studies emphasizing the effectiveness of ensemble learning in customer churn prediction and retail analytics [8, 9]. Ensemble methods are particularly effective in datasets characterized by heterogeneous purchasing behaviors because different base learners capture different dimensions of customer activity, spending patterns, and transaction regularity. The strong performance of the Stacking model in the present study also supports the argument that integrated learning frameworks provide more stable decision boundaries and greater adaptability to complex customer datasets [7].

The results also revealed that Logistic Regression performed remarkably well despite being a classical statistical classification algorithm. This finding suggests that customer churn behavior in retail settings still contains identifiable linear or semi-linear relationships among RFM indicators and churn probability. The high recall obtained by Logistic Regression indicates that the model was highly effective in identifying churn-prone customers, even though its precision was lower than some ensemble models. This finding aligns with previous studies reporting that Logistic Regression remains a competitive and interpretable baseline for churn prediction tasks, especially when behavioral indicators such as Recency, Frequency, and Monetary value are used as explanatory variables [6, 10]. From a managerial perspective, the interpretability of Logistic Regression is important because it allows decision-makers to understand the relationship between customer behavior and churn probability rather than relying solely on black-box predictions.

Naïve Bayes also demonstrated relatively strong performance in the current study, ranking among the top-performing algorithms. The acceptable performance of this algorithm may be attributed to the probabilistic structure of customer purchase behavior, where conditional independence assumptions can still provide useful approximations for predicting churn. Although Naïve Bayes is generally considered a relatively simple classification method, previous research has shown that it can achieve strong performance in customer analytics when transaction-based features are carefully engineered [18]. The present findings confirm that even relatively

lightweight probabilistic algorithms can perform effectively in retail churn prediction when supported by meaningful behavioral indicators such as RFM metrics.

The Support Vector Machine model also produced satisfactory results, indicating its ability to identify complex decision boundaries within customer transaction data. SVM models are particularly useful in classification problems involving multidimensional behavioral features because they maximize the margin between customer classes and can capture nonlinear relationships through kernel functions. The current results are aligned with earlier comparative studies that identified Support Vector Machines as robust classifiers in churn prediction applications [6]. However, despite its strong predictive capability, the SVM model did not outperform the Stacking classifier, suggesting that combining multiple models remains more effective than relying on a single sophisticated learner in highly heterogeneous retail environments.

The Decision Tree and AdaBoost models showed moderate predictive performance in the present study. Decision Trees are generally valued for their interpretability and rule-based structure, allowing managers to understand customer classification pathways in practical terms. However, the moderate results obtained in this study indicate that single-tree structures may not fully capture the complexity of retail customer behavior. AdaBoost, which iteratively improves weak learners, demonstrated better performance than several classical models but still remained below the Stacking classifier. These findings support previous research suggesting that boosting approaches improve classification performance but may still be sensitive to noisy or imbalanced datasets [8]. In retail environments characterized by highly dynamic purchasing behavior, ensemble strategies that combine heterogeneous models appear more capable of handling customer diversity.

The Bagging model produced an interesting pattern of results because it achieved exceptionally high recall but comparatively low precision and accuracy. This suggests that the model was highly sensitive to identifying churners but generated a considerable number of false positives. Such behavior is particularly important in practical retail settings because excessive false-positive predictions may lead to inefficient allocation of retention resources and unnecessary promotional expenditures. Nonetheless, the high recall obtained by Bagging may still be useful in contexts where missing a churn-prone customer is more costly than incorrectly classifying a loyal customer as at risk. Previous research in customer churn analytics has emphasized that the selection of an optimal model depends not only on overall accuracy but also on the specific managerial objectives of retention campaigns [17]. Therefore, although Bagging did not achieve the highest overall rank, its sensitivity characteristics may still offer practical value in targeted retention strategies.

The Neural Network model achieved acceptable but not leading performance in the present study. This finding suggests that deep or neural architectures do not necessarily outperform simpler ensemble or statistical methods in all retail churn scenarios. Although neural networks are capable of modeling nonlinear relationships and hidden behavioral patterns, their effectiveness depends heavily on dataset size, architecture design, feature representation, and parameter tuning. Previous studies on retail purchasing behavior and churn prediction have shown that neural and deep learning models can be highly effective in sequential and large-scale datasets [11, 12]. However, the current results indicate that in structured RFM-based retail datasets, ensemble classifiers such as Stacking may provide more stable and interpretable outcomes than standalone neural architectures. This finding is also consistent with studies emphasizing that deep learning models do not always guarantee superior performance when compared with optimized ensemble learners [13].

The K-Nearest Neighbors and Linear Discriminant Analysis algorithms demonstrated the weakest performance among the evaluated models. The relatively low effectiveness of KNN may be associated with the high

dimensionality and behavioral heterogeneity of retail transaction data, where distance-based methods struggle to identify stable neighborhood structures. Likewise, the weak performance of Linear Discriminant Analysis suggests that customer churn behavior cannot be adequately represented through simple linear separability assumptions. These findings support the broader literature indicating that simpler distance-based or linear discriminant approaches are often insufficient for capturing the complexity of retail customer behavior [2].

Beyond classification performance, the association rule analysis revealed meaningful relationships among frequently co-purchased products. The extracted rules indicated that several product combinations repeatedly appeared together in the shopping baskets of valuable non-churning customers. This finding suggests that loyal customers tend to exhibit stable and repetitive basket structures, reflecting predictable purchasing habits and complementary product preferences. The identified relationships among products provide valuable insight into customer purchasing behavior and support the argument that customer retention is closely linked to purchase consistency and basket composition. Previous studies on customer behavior mining have similarly emphasized that combining classification techniques with association rule mining can reveal hidden behavioral structures within customer transactions [16]. From a retail management perspective, these findings imply that frequently co-purchased items can be strategically utilized in cross-selling campaigns, bundle promotions, shelf arrangement optimization, and recommendation systems.

The network representation of association rules further highlighted the importance of several central products that participated in a large number of purchasing relationships. The strong Lift-based associations observed among certain items indicate that customer purchases are not independent but instead reflect interconnected consumption patterns. This finding aligns with studies emphasizing the importance of understanding complementary product relationships in retail data mining and customer segmentation [3]. Such product-level insights are highly valuable because they allow retailers to design targeted promotions that increase basket size while simultaneously improving customer satisfaction and retention.

The clustering analysis based on the K-Means algorithm also generated meaningful insights into customer segmentation. The results showed that customers could be categorized into three major clusters, including loyal high-value customers, moderately active potential customers, and customers at risk of churn. This finding demonstrates the practical value of integrating RFM indicators with clustering techniques in retail analytics. The loyal customer cluster exhibited lower Recency values and higher Monetary scores, indicating recent and financially valuable purchases, whereas churn-prone customers showed longer purchase intervals and lower spending levels. These findings are consistent with previous studies reporting that RFM-based segmentation effectively differentiates customer groups according to value and behavioral intensity [3, 4]. The overlap between the association rule patterns and the loyal customer clusters also confirms that profitable customers tend to demonstrate repetitive and structured purchasing behavior.

The present study contributes to the literature by integrating multiple analytical perspectives, including RFM modeling, machine learning classification, association rule mining, and clustering analysis within a unified retail churn prediction framework. Unlike studies focusing exclusively on predictive accuracy, this research provides a broader interpretation of customer behavior and demonstrates how predictive and descriptive analytics can complement one another in retail decision-making. The findings also reinforce the importance of using actual transaction data rather than hypothetical or simulated datasets, because real-world retail data contain behavioral complexities that strongly influence model performance. Moreover, the use of multiple evaluation metrics,

including Accuracy, Precision, Recall, and F1-score, provided a comprehensive assessment of classifier effectiveness and avoided overreliance on a single performance indicator.

The results of this study also have important strategic implications for retail management in omni-channel and data-driven environments. Contemporary retail systems increasingly rely on predictive analytics to optimize customer retention strategies, personalize marketing campaigns, and improve customer lifetime value [1]. In this context, the ability to identify churn-prone customers using RFM-based machine learning models can significantly improve managerial responsiveness and customer relationship management effectiveness. Furthermore, the findings support broader trends in customer analytics emphasizing the integration of segmentation, classification, and predictive modeling into comprehensive retail intelligence systems [14, 19]. As retail competition intensifies and customer acquisition costs continue to rise, predictive churn modeling is likely to become an increasingly essential component of strategic retail management.

One of the limitations of the present study is that the dataset was extracted from a single retail store branch, which may reduce the generalizability of the findings to other retail environments, geographical regions, or customer populations. Additionally, the study relied primarily on transactional and behavioral variables represented through RFM indicators and did not include demographic, psychographic, or attitudinal variables that may further improve churn prediction performance. Another limitation concerns the static nature of the dataset, as customer behavior was analyzed within a fixed time interval rather than through a fully longitudinal or real-time framework. Moreover, although several machine learning algorithms were compared, extensive hyperparameter optimization and advanced deep learning architectures were beyond the scope of the current research.

Future studies should investigate customer churn prediction using larger multi-store or multi-channel retail datasets to improve model generalizability and robustness. Researchers may also incorporate additional behavioral, demographic, and online interaction variables to enrich the predictive feature space and better capture complex customer behavior. Future research can further examine hybrid and deep learning frameworks, including recurrent neural networks, transformer-based architectures, and sequential learning models, especially in dynamic retail environments where customer behavior evolves over time. In addition, future studies may compare real-time churn prediction systems with static classification models and evaluate the economic impact of churn prediction on retention campaign effectiveness and customer lifetime value.

From a practical perspective, retail managers can use the findings of this study to develop more targeted and data-driven customer retention strategies. The identified high-performing models, especially the Stacking classifier, can be integrated into retail decision-support systems to detect customers at risk of churn before disengagement occurs. The association rule findings can assist managers in designing effective cross-selling campaigns, product bundles, and optimized store layouts, while clustering results can support customer segmentation and personalized loyalty programs. Retail firms may also use RFM-based predictive systems to prioritize marketing resources toward high-value and at-risk customer segments, thereby improving customer retention efficiency and long-term profitability.

Authors' Contributions

Authors equally contributed to this article.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all participants who participate in this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

- [1] M. H. Fathali, Kambiz, R. Zaboli, and M. Khoun Siavash, "A phenomenological study of omni-channel intensity as a key factor influencing omni-channel shopping value," *Journal of Marketing Management*, vol. 20, no. 3, pp. 1-14, 2025.
- [2] D. Chandrakala, "A survey on customer churn prediction using machine learning techniques," *International Journal of Computer Applications*, vol. 154, no. 10, 2016, doi: 10.5120/ijca2016912237.
- [3] M. A. Rahim, "RFM-based repurchase behavior for customer classification and segmentation," *Journal of Retailing and Consumer Services*, vol. 61, p. 102566, 2021, doi: 10.1016/j.jretconser.2021.102566.
- [4] I. Lewaaelhamd, "Customer segmentation using machine learning model: An application of RFM analysis," *Journal of Data Science and Intelligent Systems*, vol. 2, no. 1, pp. 29-36, 2024, doi: 10.47852/bonviewJDSIS32021293.
- [5] A. Harish and C. Malathy, "Evaluative study of cluster based customer churn prediction against conventional RFM based churn model," in *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, 2023: IEEE.
- [6] S. Khodabandehlou and M. Zivari Rahman, "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior," *Journal of Systems and Information Technology*, vol. 19, no. 1-2, pp. 65-93, 2017, doi: 10.1108/JSIT-10-2016-0061.
- [7] X. Guoen and H. Qingzhe, "The Research of Online Shopping Customer Churn Prediction Based on Integrated Learning," in *Proceedings of the 2018 International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018)*, 2018: Atlantis Press.
- [8] M. Bogaert and L. Delaere, "Ensemble methods in customer churn prediction: A comparative analysis of the state-of-the-art," *Mathematics*, vol. 11, no. 5, p. 1137, 2023, doi: 10.3390/math11051137.
- [9] S. Awasthi, "Customer churn prediction on e-commerce data using stacking classifier," *Authorea Preprints*, 2022, doi: 10.36227/techrxiv.20291694.
- [10] B. Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," *International Journal of Intelligent Networks*, vol. 4, pp. 145-154, 2023, doi: 10.1016/j.ijin.2023.05.005.
- [11] A. Verma, "Consumer behaviour in retail: Next logical purchase using deep neural network," *arXiv preprint arXiv:2010.06952*, 2020.
- [12] J. P. Equihua, "Modelling customer churn for the retail industry in a deep learning based sequential framework," *arXiv preprint arXiv:2304.00575*, 2023.
- [13] M. Mahdi and M. Jabbari, "Predicting customer churn in the fast-moving consumer goods segment of the retail industry using deep learning," *Mathematics and Computational Sciences*, vol. 5, no. 3, pp. 58-79, 2024.
- [14] S. K. Wagh, "Customer churn prediction in telecom sector using machine learning techniques," *Results in Control and Optimization*, vol. 14, p. 100342, 2024, doi: 10.1016/j.rico.2023.100342.
- [15] J. B. Brito, "A framework to improve churn prediction performance in retail banking," *Financial Innovation*, vol. 10, no. 1, p. 17, 2024, doi: 10.1186/s40854-023-00558-3.
- [16] F. Abdi and S. Abolmakarem, "Customer Behavior Mining Framework (CBMF) using clustering and classification techniques," *Journal of Industrial Engineering International*, vol. 15, no. Suppl 1, pp. 1-18, 2019, doi: 10.1007/s40092-018-0285-3.

- [17] H. Casteran, L. Meyer-Waarden, and W. Reinartz, "Modeling customer lifetime value, retention, and churn," in *Handbook of Market Research*: Springer, 2017, pp. 1-33.
- [18] Z. Liang, "Predict customer churn based on machine learning algorithms," *HBEM*, vol. 10, pp. 270-275, 2023, doi: 10.54097/hbem.v10i.8051.
- [19] S. N. Gunesen, "Customer churn prediction in FMCG sector using machine learning applications," in *IFIP International Workshop on Artificial Intelligence for Knowledge Management*, 2021: Springer, doi: 10.1007/978-3-030-80847-1_6.