Article history:
Received 22 December 2025
Revised 23 April 2026
Accepted 30 April 2026
Published online 01 May 2026

Journal of Resource Management and Decision Engineering

Volume 5, Issue 3, pp 1-21

Simultaneous Use of Clustering, Classification, and Association Rules for Customer Churn Analysis in Retail Stores

Mohammad. Ghasemi Zadeh[1], Seyed Abdollah Amin. Mousavi[2*], Mohammad. Maleki Nia[3]

[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

*** Corresponding author email address**: saa.mousavi@iau.ac.ir

Article Info

Article type:
Original Research

How to cite this article:
Ghasemi Zadeh, M., Mousavi, S. A. A., & Maleki Nia, M. (2026). Simultaneous Use of Clustering, Classification, and Association Rules for Customer Churn Analysis in Retail Stores. *Journal of Resource Management and Decision Engineering*, *5*(3), 1-21.
https://doi.org/10.61838/kman.jrmde.303

A B S T R A C T

Customer churn analysis plays a critical role in sustaining profitability in competitive retail environments. This study proposes an integrated data mining framework that combines clustering, classification, and association rule mining within an enhanced RFM-based model, referred to as RFFML, which extends the traditional Recency, Frequency, and Monetary indicators by incorporating Frequency Items and Length of Interaction variables. Using large-scale real transactional data from a retail store, customer segmentation is first performed through clustering to identify distinct behavioral profiles. Subsequently, several classification models are evaluated for churn prediction, with an ensemble-based Stacking classifier achieving the highest predictive performance. In parallel, association rule mining is employed to uncover frequent product combinations associated with loyal and churned customer segments. The main contribution of this work lies in the analytical integration of these complementary techniques within a single unified framework, enabling both accurate churn prediction and interpretable behavioral insights. Experimental results demonstrate that the proposed approach outperforms individual classifiers while providing actionable knowledge to support targeted retention strategies in the retail sector.

Keywords: *Customer Churn Prediction, Retail Data Mining, RFM Model Enhancement (RFFML), Clustering–Classification–Association Rules, Stacking Ensemble Method*

1. Introduction

Customer retention has become one of the most strategic priorities for organizations operating in highly competitive retail environments, where increasing market saturation, digital transformation, and changing consumer expectations continuously intensify competition among firms. Retail companies are increasingly shifting their focus from transactional sales toward long-term customer relationship management, recognizing that retaining existing customers is significantly more cost-effective than acquiring new ones (Casteran et al., 2017; Hughes, 1994). In this context, customer churn, defined as the discontinuation or significant reduction of purchasing behavior by customers, represents a major threat to organizational profitability, sustainability, and market share. Churn not only leads to direct revenue loss but also increases marketing expenditures associated with customer acquisition and promotional campaigns (Lalwani et al., 2022; Malhotra et al., 2025). Consequently, identifying customers at risk of churn and implementing preventive retention strategies have become critical managerial objectives across customer-oriented industries.

The rapid growth of big data technologies and computational intelligence has enabled organizations to leverage transactional and behavioral data for predictive decision-making. Machine learning and data mining techniques have emerged as powerful analytical tools capable of identifying hidden behavioral patterns, predicting customer actions, and supporting strategic planning in retail environments (Mining, 2006; Pasquadibisceglie et al., 2024). These technologies allow firms to move beyond traditional descriptive analysis toward predictive and prescriptive analytics capable of generating actionable insights for customer retention. Recent studies have demonstrated that predictive analytics systems can substantially improve the personalization of customer services, optimize promotional targeting, and increase customer loyalty through accurate prediction of behavioral tendencies (Bhargava et al., 2024; Vemulapalli, 2024). In modern retail ecosystems characterized by large-scale transactional data and dynamic purchasing behavior, the integration of machine learning into customer relationship management has therefore become both a technological necessity and a strategic advantage.

Customer churn prediction has attracted significant scholarly attention during the past decade, particularly in industries such as telecommunications, banking, insurance, and e-commerce. Researchers have increasingly employed machine learning algorithms including Decision Trees, Logistic Regression, Support Vector Machines, Artificial Neural Networks, Random Forests, and ensemble learning approaches to improve churn prediction accuracy (Gaur & Dubey, 2018; Prabadevi et al., 2023). These models are capable of learning complex nonlinear relationships between customer attributes and churn behavior, thereby outperforming traditional statistical methods in many applications. For example, Lalwani et al. demonstrated the effectiveness of machine learning classification systems in predicting customer churn using transactional and behavioral variables (Lalwani et al., 2022). Similarly, Malhotra et al. emphasized the growing applicability of predictive machine learning frameworks for churn analysis in highly competitive service industries (Malhotra et al., 2025). Ensemble-based approaches, particularly boosting and stacking methods, have also shown considerable promise in improving predictive robustness and reducing classification errors by combining the strengths of multiple classifiers (Khandelwal & Sakalle, 2025).

Despite these advances, most existing churn prediction studies remain concentrated in sectors such as telecommunications and online services, while the retail sector has received comparatively less analytical attention (Matuszelanski & Kopczewska, 2022; Mendez, 2020). Retail environments present unique analytical challenges because customer behavior is highly dynamic, purchase frequency varies substantially among customers, and purchasing decisions are influenced by a broad range of behavioral, demographic, and situational factors. Furthermore, retail datasets are often characterized by high dimensionality, sparse transaction patterns, and large numbers of product combinations, making predictive analysis substantially more complex. Consequently, the direct transferability of churn models developed in telecommunications or banking to retail contexts remains limited. This gap highlights the need for specialized retail-oriented frameworks capable of simultaneously analyzing customer purchasing intensity, loyalty patterns, and product associations.

One of the foundational analytical approaches used in customer behavior analysis is the RFM model, originally proposed within database marketing literature (Hughes, 1994). The RFM framework evaluates customer value and behavioral tendencies according to three dimensions: Recency of purchase, Frequency of transactions, and Monetary value of purchases. Customers with recent purchases, frequent transactions, and high spending are generally considered more loyal and profitable, whereas customers exhibiting declining RFM values are viewed as potential churn candidates. Numerous studies have confirmed the predictive capability of RFM variables for customer segmentation, loyalty analysis, and churn prediction (Bagul et al., 2021; Rahim et al., 2021). Rahim et al. demonstrated that RFM indicators significantly improve the classification of customer groups and repurchase behavior in retail settings (Rahim et al., 2021). Similarly, Bagul et al. integrated RFM analysis with K-Means clustering to identify customer segments exhibiting different churn risks in retail environments (Bagul et al., 2021). Although these findings support the effectiveness of RFM-based analysis, many existing studies employ the classical RFM framework without extending its variables or integrating it comprehensively with advanced machine learning techniques.

In addition to RFM analysis, clustering methods have become increasingly important in customer behavior mining because they enable the identification of homogeneous customer segments based on purchasing characteristics and behavioral similarities. Clustering algorithms such as K-Means provide unsupervised mechanisms for identifying latent structures within customer data, thereby supporting targeted marketing and personalized retention strategies (Abdi & Abolmakarem, 2019). Abdi and Abolmakarem proposed a hybrid customer behavior mining framework that combined clustering and classification techniques to improve churn prediction performance (Abdi & Abolmakarem, 2019). Their findings indicated that customer segmentation prior to classification substantially enhances the identification of high-risk customer groups. Similarly, Matuszelanski and Kopczewska employed spatial and machine learning approaches to investigate customer churn in retail e-commerce settings and demonstrated the importance of segmentation-based analysis in understanding purchasing behavior (Matuszelanski & Kopczewska, 2022). Nevertheless, most prior studies have used clustering merely as a preprocessing tool rather than integrating it strategically with classification and association rule mining within a unified analytical architecture.

Classification models constitute another major component of churn prediction systems. Supervised machine learning algorithms aim to predict whether a customer belongs to a churn or non-churn category based on historical behavioral attributes. Numerous studies have compared the performance of different classifiers in churn analysis contexts (Gaur & Dubey, 2018; Prabadevi et al., 2023). Gaur and Dubey evaluated multiple machine learning methods for churn prediction in telecommunications and found that ensemble learning methods often outperform standalone algorithms in predictive accuracy (Gaur & Dubey, 2018). Similarly, Schaeffer and Sanchez highlighted the capability of machine learning models to forecast customer retention using transactional and behavioral data (Schaeffer & Sanchez, 2020). Recent developments in explainable artificial intelligence have further emphasized the need for interpretable prediction systems capable of balancing predictive power with managerial usability (Pasquadibisceglie et al., 2024; Valente et al., 2022). Pasquadibisceglie et al., for instance, introduced an explainable process prediction mining approach for customer churn prediction in evolving retail data environments, demonstrating that explainability significantly enhances managerial trust in machine learning outputs (Pasquadibisceglie et al., 2024).

Another important but relatively underutilized analytical technique in churn analysis is association rule mining. Association rules identify frequent co-purchase relationships among products and reveal hidden patterns in customer shopping baskets. Such information is particularly valuable in retail environments because it supports cross-selling, personalized recommendation systems, promotional bundling, and customer engagement strategies. Although association rule mining has been widely applied in market basket analysis, its integration with churn prediction frameworks remains limited (Mining, 2006). Existing research often treats association rule analysis separately from classification and clustering processes, thereby missing opportunities to generate comprehensive behavioral insights. Integrating association rule mining with churn prediction could enable retailers not only to identify customers at risk of churn but also to determine the specific purchasing patterns associated with loyal or disengaged customers.

Recent developments in artificial intelligence and deep learning have further expanded the analytical capabilities available for customer behavior prediction. Verma employed deep neural networks to predict future purchasing behavior in online retail contexts and demonstrated superior predictive performance compared with traditional regression-based methods (Verma, 2020). Similarly, Wu et al. emphasized the effectiveness of advanced feature learning approaches in improving predictive modeling accuracy in complex behavioral datasets (Wu et al., 2024).

Chen et al. and Cheng et al. also highlighted the increasing applicability of hybrid and coupled machine learning models in analyzing complex multidimensional systems (Chen et al., 2023; Cheng et al., 2023). These studies collectively suggest that hybrid analytical architectures combining multiple machine learning paradigms are more capable of capturing the multidimensional nature of customer behavior than isolated algorithms.

The growing adoption of predictive analytics in customer relationship management has also intensified interest in personalization and customer-centric decision systems. Bhargava et al. demonstrated that predictive analytics substantially enhances service personalization and customer satisfaction in e-commerce systems (Bhargava et al., 2024). Similarly, Valente et al. argued that reliability, interpretability, and personalization are essential characteristics of effective machine learning decision-support systems (Valente et al., 2022). In churn management contexts, personalized interventions based on behavioral segmentation and predictive analytics can significantly improve customer retention outcomes. Vemulapalli further emphasized that AI-driven predictive models can proactively identify customer dissatisfaction patterns and support strategic retention initiatives before customers disengage completely (Vemulapalli, 2024). These findings reinforce the strategic importance of integrating behavioral analytics, machine learning, and personalized marketing approaches within retail management systems.

Despite the substantial growth of research in churn prediction and customer analytics, several important gaps remain evident in the literature. First, most studies focus on isolated analytical techniques, such as classification or clustering, rather than employing integrated frameworks that simultaneously combine clustering, classification, and association rule mining. Second, relatively few studies have concentrated specifically on large-scale retail transaction datasets despite the distinctive behavioral complexity of retail customers (Agbemadon et al., 2022; Mendez, 2020). Third, although the RFM model has been widely recognized as an effective behavioral framework, its variables are often used in their traditional form without extension or integration into comprehensive machine learning pipelines. Fourth, many studies prioritize predictive accuracy while neglecting behavioral interpretability and practical managerial applicability. Finally, the role of enhanced behavioral indicators and ensemble learning approaches in improving retail churn analysis remains insufficiently explored.

To address these research gaps, the present study proposes an integrated analytical framework for customer churn analysis in retail environments by simultaneously combining clustering, classification, and association rule mining within an enhanced RFM-based model. The proposed framework extends the traditional RFM structure through the incorporation of additional behavioral indicators and utilizes ensemble learning techniques to improve predictive performance while maintaining interpretability and operational applicability. Using large-scale real transactional data from a retail store, this study aims to develop a comprehensive customer churn analysis system capable of generating both accurate churn predictions and actionable behavioral insights for strategic retail decision-making.

2. Methods and Materials

This section outlines the research methodology framework for evaluating and predicting customer churn in the retail sector. The overall approach is based on the standard CRISP-DM framework, incorporating real transactional data, feature engineering, and the use of advanced machine learning techniques. The methodology is organized into three main stages: (1) data preparation and research design, (2) the proposed hybrid modeling approach integrating clustering, association rule mining, and classification, and (3) model validation and final evaluation.

This study is an applied research project based on actual transaction data from a retail store. The overall approach is designed in accordance with the CRISP-DM framework, encompassing the stages of data understanding, business understanding, modeling, data preparation, evaluation, and deployment of results.

Dataset: The dataset includes information such as purchase date and amount, sales transactions, purchase frequency, and other demographic and behavioral attributes of customers. RFM indicators are employed as the key features. The target variable is the customer churn status, labeled based on the absence of purchases within a specified time period.

Data Preparation: This stage involves data cleaning (removing duplicate records and those with noisy or missing values), data normalization, feature selection and engineering, encoding of categorical variables, and class balancing using techniques such as SMOTE, to ensure the dataset is ready for modeling.

The model proposed in this study combines three primary data mining techniques:

1. Clustering:

- Applying the K-Means algorithm to segment customers based on RFM metrics and other attributes.

- Determining the optimal number of clusters using the Silhouette Score and Elbow Method factors.

2. Classification:

- Implementing algorithms such as Random Forest, Decision Tree, Gradient Boosting, and Logistic Regression.

- Selecting the final model based on evaluation metrics and feature importance analysis.

3. Association Rules:

- Association rule mining was conducted exclusively on the training dataset after the train–test split to prevent information leakage.

- Identifying frequent purchasing patterns among churn and non-churn customers within the training data using the Apriori algorithm, without utilizing churn labels from the test set.

- Evaluating the extracted rules according to Confidence, Support, and Lift measures.

No transaction records from the test dataset were used during the association rule extraction stage, ensuring unbiased pattern discovery and real-world applicability of the results. Finally, the outcomes of these three modules are integrated to derive comprehensive behavioral patterns, enabling both causal analysis and churn prediction.

The dataset was randomly divided into training (70%) and testing (30%) subsets. All classification models were trained exclusively on the training data, and hyperparameter tuning was performed using internal validation within the training subset. The Stacking ensemble was constructed by combining multiple base learners, including Decision Tree, KNN, SVM, AdaBoost, and Bagging, with a meta-learner trained solely on the training data.

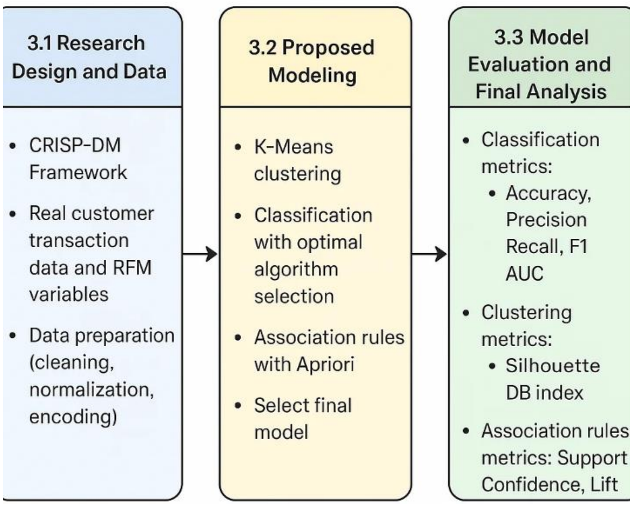
The evaluation criteria applied in this study are as follows:

- **Classification:** Accuracy, Precision, Recall, F1-Score, ROC Curve, and AUC.
- **Clustering:** Silhouette Coefficient and Davies–Bouldin Index, accompanied by qualitative interpretation of each cluster.
- **Association Rules:** Support, Confidence, and Lift.

In this study, after individually assessing each method, the best-performing model or combination of models is selected based on separation capability, accuracy, interpretability, and operational value. The resulting model can assist store managers in accurately segmenting customers and, through targeted strategies, reducing churn rates. Figure 1 presents the overall methodology process, including data preparation, the integration of research design with the proposed modeling framework, and the subsequent stages of model examination and final evaluation.

Figure 1

Methodology Workflow for Customer Churn Analysis in Retail



This evaluation protocol ensures reproducibility, prevents information leakage, and enables an unbiased assessment of model performance.

3. Findings and Results

In this section, the outputs generated by the proposed model are presented and analyzed according to the standard CRISP-DM process. The analyses were conducted using real-world transactional data from a retail store and were based on the developed RFFML indicators. The main objective was to transform raw transaction records into reliable analytical patterns for identifying non-churn and churn customers, recognizing high-value customers within each group, and extracting interpretable purchasing patterns that could support managerial decision-making. The stages of analysis included data preparation, RFFML feature construction, customer clustering, churn classification, algorithm comparison, and association rule mining.

The dataset was obtained from the “Seraj Branch 1” retail store and included approximately 541,000 transaction records over the observation period. The dataset contained 13 primary attributes related to purchasing behavior, including product barcode, season, month, day of the week, sales amount, customer identifier, number of days since the last purchase, number of items sold, and number of sales invoices. Among these attributes, five indicators formed the core of the proposed RFFML model: Recency (R), Frequency Transactions (F_1), Frequency Items (F_2), Monetary value (M), and Length of Interaction (L). The traditional RFM model was therefore extended by adding F_2, which captures the total volume of purchased items, and L, which represents the duration of the customer–store relationship. These additional indicators enabled the model to capture purchasing intensity and customer relationship stability more accurately than the classical RFM structure.

Figure 2

Conceptual framework of the RFFML model with the Length of Interaction (L) index

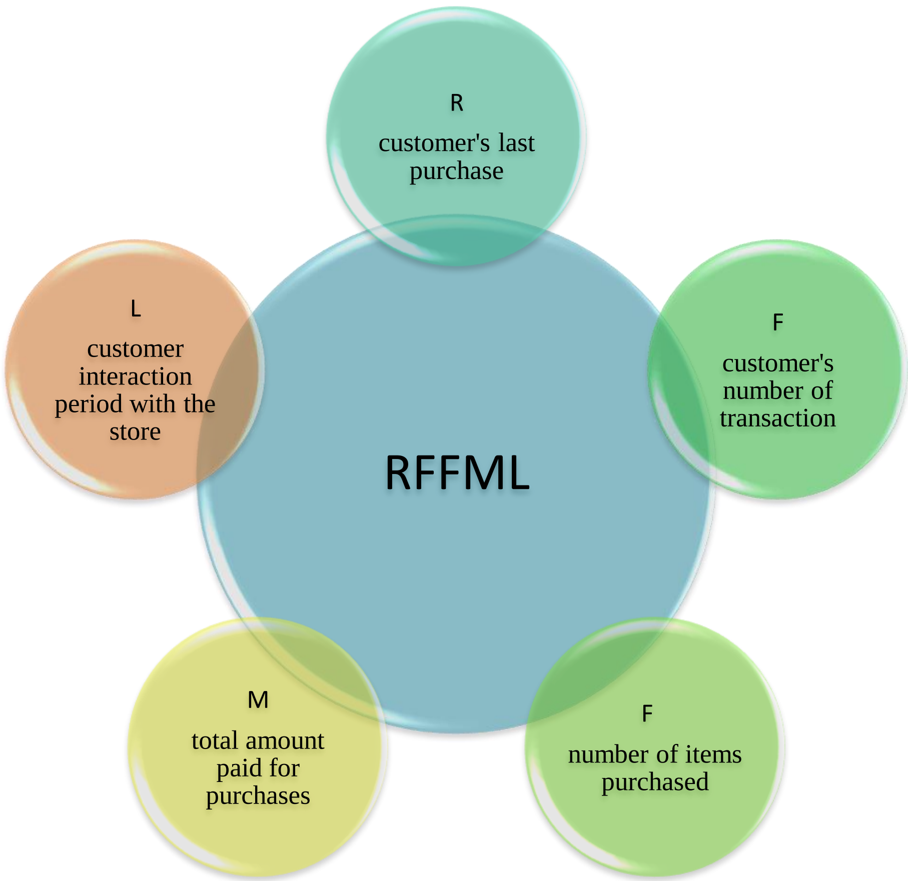


Table 1

Structure of transactional data before preprocessing in customer churn analysis

Store Name	Item Name	Product Barcode	Customer ID	Season	Month	Day of the Week	Day	Net Sales Amount after Returns (Rials)	Last Sales Date	Number of Days Elapsed Since Last Sale	Total Number of Items Sold	Number of Sales Invoices
Seraj Branch 1	Saie Low-Absorption Frying Oil 810g	6260295101705	9010222018	Summer	August	Thursday	20	613,040	14010520	-	1	1
Seraj Branch 1	Softlan Complete & Shiny Gold Machine Washing Powder 500g Pack of 24	6260010500394	9010222018	Summer	August	Thursday	20	153,546	14010520	-	1	1
Seraj Branch 1	Pegah Domestic Butter 25g	6260707400587	9010222018	Summer	August	Thursday	20	59,037	14010520	-	1	1
Seraj Branch 1	Schon Fruity Mix Shampoo 400ml	6260482526373	9010222018	Summer	August	Thursday	20	436,055	14010520	-	1	1
Seraj Branch 1	Camelion AA Alkaline Batteries Pack of 2	4260033151032	9010222018	Summer	August	Thursday	20	90,826	14010520	-	1	1
Seraj Branch 1	Signal Anti-Cavity Toothpaste 100 ml	6260526418176	9010222018	Summer	August	Thursday	20	234,862	14010520	-	1	1
Seraj Branch 1	Kalleh White UF Cheese 400 g	6260161529220	9010222018	Summer	August	Thursday	20	288,000	14010520	-	1	1

As shown in Table 1, the original transaction-level data contained repeated store-level information, customer identifiers, calendar attributes, sales amounts, product codes, and purchase-volume indicators. Attributes such as Store Name had identical values across all records and therefore did not provide analytical differentiation. In contrast, numerical and behavioral attributes were retained because they could be used for feature engineering and machine learning procedures. The presence of product-level and customer-level transaction records provided a suitable basis for identifying churn patterns, customer value segments, and frequent product combinations.

Data preparation was carried out before model execution to transform raw transaction records into a customer-level analytical dataset. Records with unusable, incomplete, zero, or negative sales values were identified and either removed or corrected where appropriate. Key numerical variables, including Monetary value, were normalized to prevent scale differences from distorting the outputs of clustering and classification algorithms. Categorical variables such as Month and Day of the Week were converted into numerical representations to ensure compatibility with the selected algorithms. The Pivot tool in Tableau was used to restructure the dataset so that each final row represented one customer instead of one transaction.

The RFFML indicators were calculated as follows:

$$R_i = T_{\text{reference}} - \max(t_{ij})$$

$$F_{1i} = \sum_{j=1}^{n_i} I(\text{transaction}_{ij})$$

$$F_{2i} = \sum_{j=1}^{n_i} \text{items}_{ij}$$

$$M_i = \sum_{j=1}^{n_i} \text{net sales amount}_{ij}$$

$$L_i = \max(t_{ij}) - \min(t_{ij})$$

In these formulas, R_i represents the number of days since the last purchase of customer i, F_{1i} represents the total number of transactions, F_{2i} represents the total number of purchased items, M_i represents total monetary spending, and L_i represents the length of interaction between the customer and the store. The churn label was defined using the following rule:

$$\text{Churn}_i = \begin{cases} 1, & R_i > 90 \\ 0, & R_i \leq 90 \end{cases}$$

Thus, customers with no purchase activity for more than 90 consecutive days were labeled as churn customers, while all other customers were labeled as non-churn customers. The dataset was randomly divided into training and testing subsets using a 70/30 ratio. Feature standardization was performed before modeling to ensure that all input variables were placed on comparable scales.

Figure 3

Preparation and loading of transactional data for the RFFML model

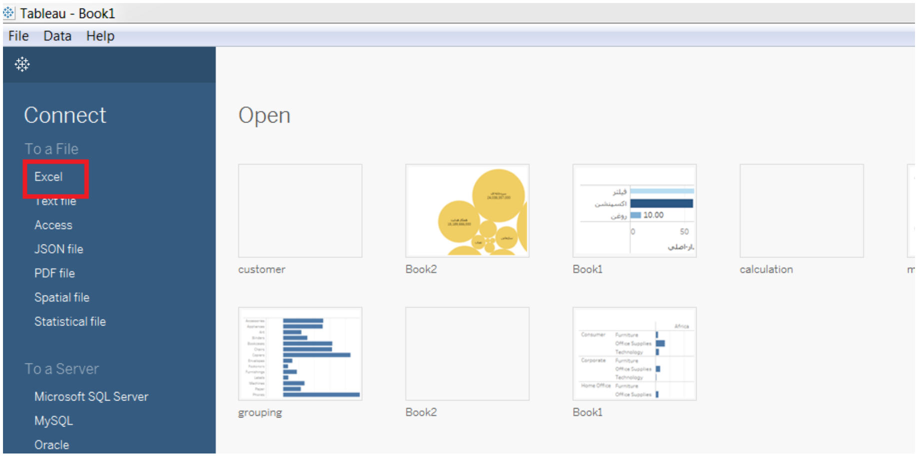


Figure 4

Management and display of transactional variables in the preprocessing stage

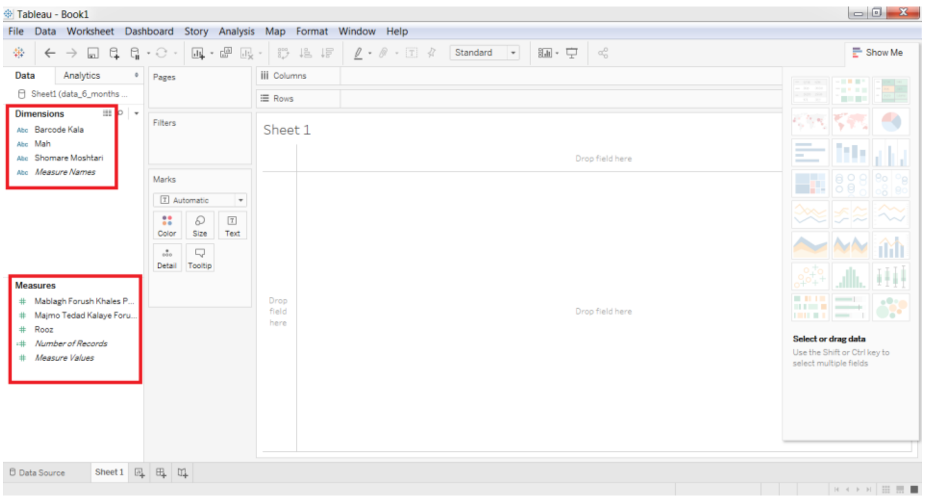


Figure 5

Configuration of transactional variables to build each customer's profile

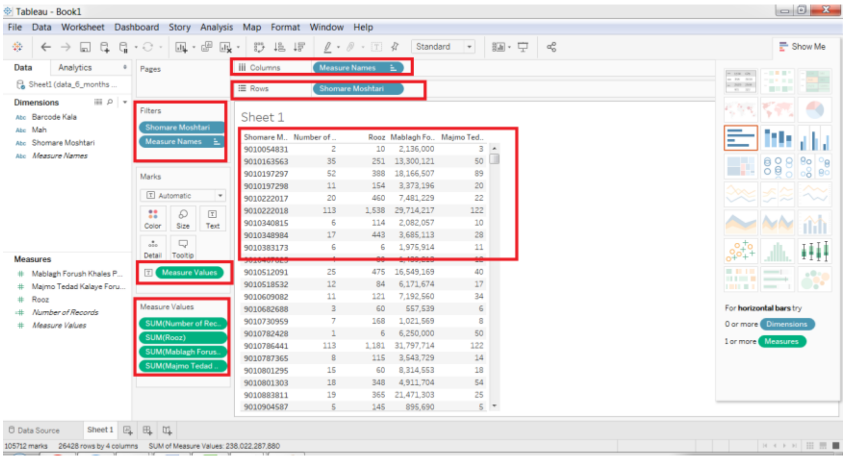


Figure 6

Method for calculating customers' most recent purchase for loyalty analysis and evaluation

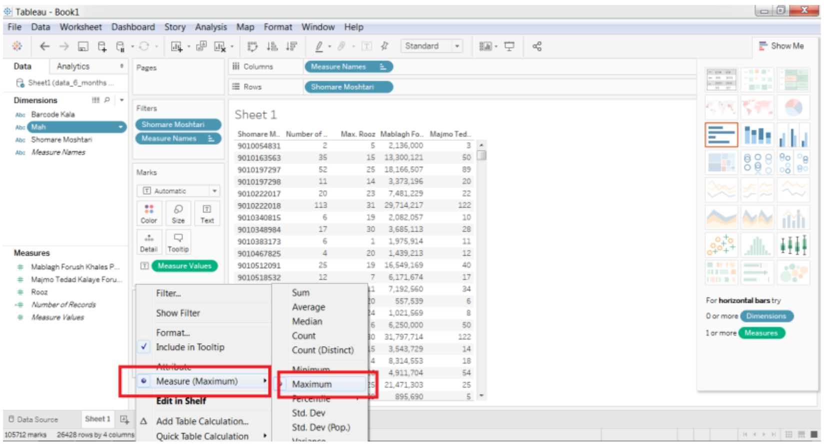


Figure 7

Final structure of customer data based on RFFML model indicators

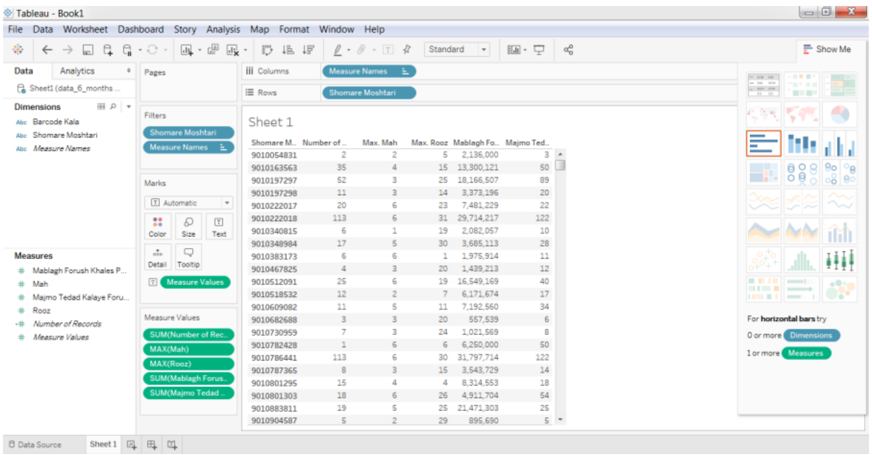
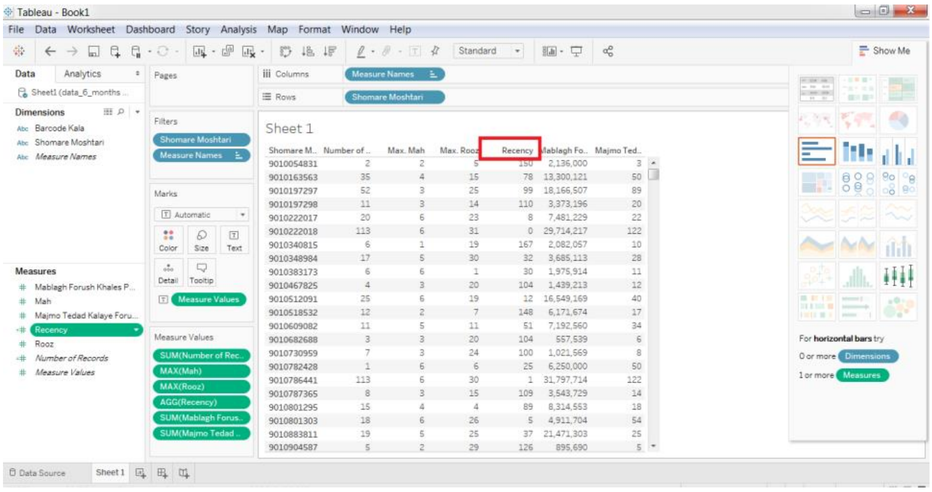


Figure 8

Data configuration and standardization before executing the RFFML model



The modeling stage transformed the prepared customer-level dataset into input structures suitable for machine learning and data mining. Three analytical procedures were implemented. First, K-Means clustering was used to segment customers according to RFFML indicators and identify behavioral groups. Second, classification algorithms were trained to predict customer churn at the individual level. Third, association rule mining was applied to detect frequent product combinations among selected customer groups. The K-Means clustering process included initialization of cluster centers, determination of the optimal number of clusters, and convergence assessment.

The K-Means model was based on the minimization of the within-cluster sum of squares:

$$WCSS = \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2$$

where C_k represents cluster k, x_i represents an individual customer observation, and μ_k represents the centroid of cluster k. The optimal value of K was determined through the Elbow Method. The analysis of the intra-cluster sum of squares indicated that $K = 4$ provided the most appropriate balance between model simplicity and reduced intra-cluster error.

Figure 9

Customer segmentation based on RFFML model features

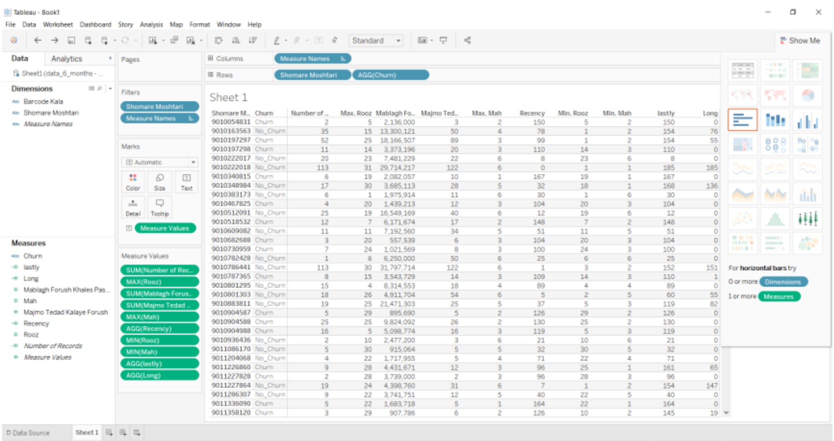


Figure 10

Analysis of feature importance in predicting customer churn

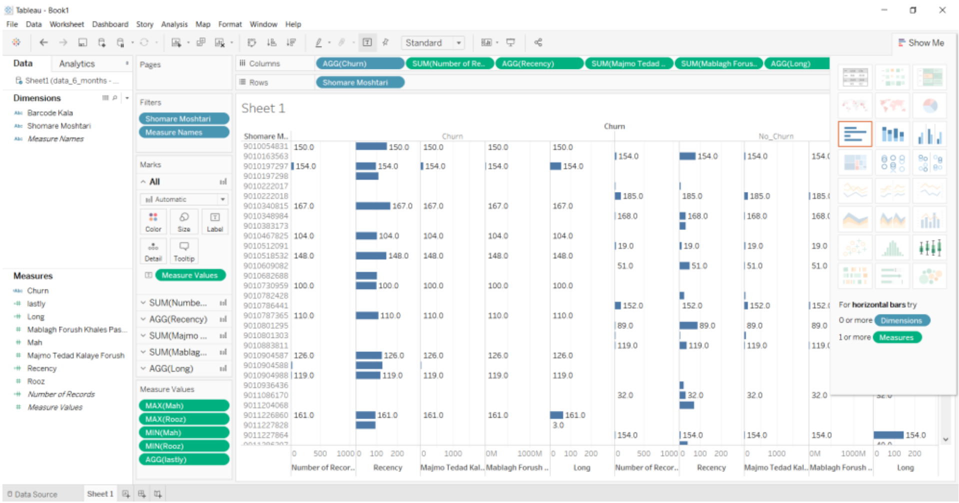


Figure 11

Identification of the optimal K value in the K-Means algorithm for store customers

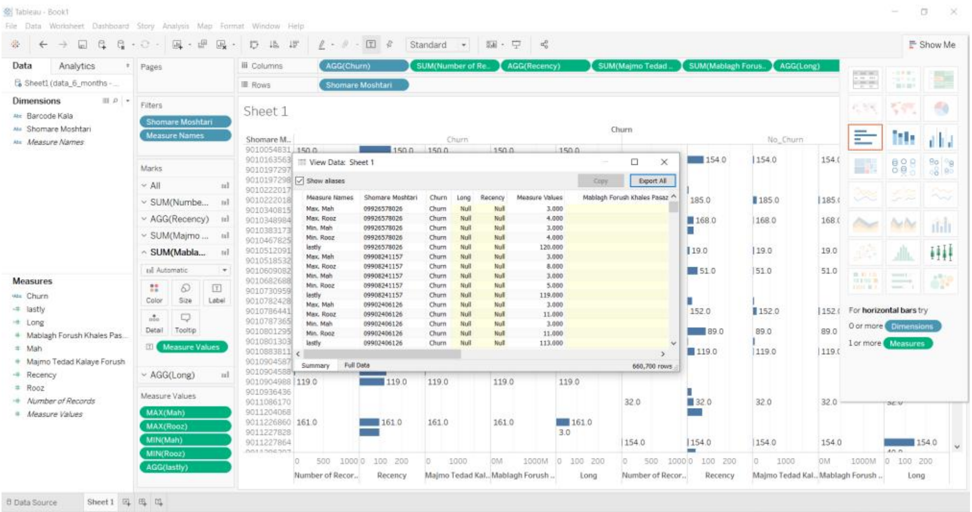


Table 2

Segmentation of customers into behavioral clusters based on the RFFML model

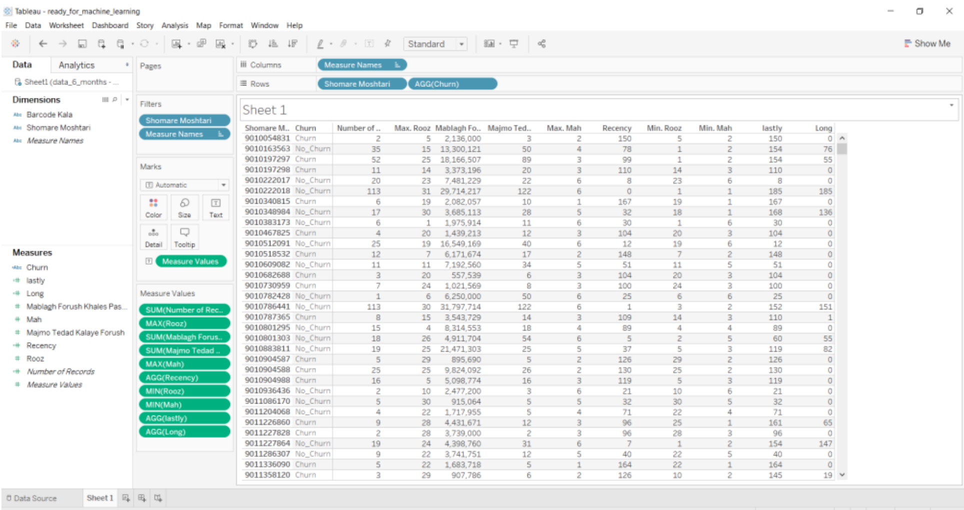
Feature	Cluster 1	Cluster 2
Last Transaction / Recency (R)	27.744	85.006
Number of Records / Transactions (F_1)	82.996	15.343
Number of Items (F_2)	133.986	23.234
Sales Amount / Monetary Value (M)	43,299,717.192	6,196,814.110
Length of Interaction (L)	117.383	29.950

The clustering results revealed clear behavioral differences between customer groups. Cluster 1 showed a low Recency value, high transaction frequency, high item frequency, high monetary value, and long interaction length. This cluster therefore represented loyal and high-value customers with recent purchasing behavior and strong engagement with the store. Cluster 2 showed a considerably higher Recency value and much lower values for Frequency

Transactions, Frequency Items, Monetary value, and Length of Interaction. This cluster therefore represented customers with weaker engagement and higher churn risk. Beyond these numerical results, the full cluster interpretation indicated four behavioral groups: Cluster 1 consisted of loyal customers characterized by recent purchases, frequent transactions, and high spending; Cluster 2 represented churned customers with a long interval since the last purchase and low monetary and frequency values; Cluster 3 represented potential customers who had recent purchases but moderate monetary value; and Cluster 4 represented low-value customers who showed neither high loyalty nor frequent purchasing. This segmentation provided the basis for targeted marketing strategies and subsequent interpretation of classification and association-rule outputs.

Figure 12

Visualization of clusters in the RFFML model feature space



After the clustering stage, classification models were applied to the labeled dataset. The input features included R, F_1, F_2, M, and L, and the target variable was the binary churn label. The models tested were Decision Tree, K-Nearest Neighbor, Naïve Bayes, Artificial Neural Network, Logistic Regression, Support Vector Machine, Linear Discriminant Analysis, AdaBoost, Bagging, and Stacking. The performance of classification algorithms was evaluated using Accuracy, Precision, Recall, and F-Measure. These metrics were calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F\text{-}Measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where TPindicates true positives, TNindicates true negatives, FPindicates false positives, and FNindicates false negatives. The confusion matrices and performance metrics for the three main ensemble models are presented in Table 3.

Table 3

Confusion matrices and performance metrics of the main ensemble classifiers

Algorithm	Predicted Class	True No Churn	True Churn	Accuracy	Precision	Recall	F-Measure
AdaBoost	Predicted No Churn	137	21	84.12	81.87	87.65	84.66
AdaBoost	Predicted Churn	33	149	84.12	81.87	87.65	84.66
Bagging	Predicted No Churn	78	2	72.35	64.62	98.82	78.14
Bagging	Predicted Churn	92	168	72.35	64.62	98.82	78.14
Stacking	Predicted No Churn	141	17	86.47	84.07	90.00	86.93
Stacking	Predicted Churn	29	153	86.47	84.07	90.00	86.93

The AdaBoost model achieved an accuracy of 84.12%, precision of 81.87%, recall of 87.65%, and F-Measure of 84.66%. The Bagging model produced the highest recall among the three ensemble models, with a recall value of 98.82%; however, its overall accuracy was lower at 72.35%, indicating tha

t it identified most churn customers but produced a larger number of incorrect classifications. The Stacking model achieved the strongest overall performance, with an accuracy of 86.47%, precision of 84.07%, recall of 90.00%, and F-Measure of 86.93%. These results show that the Stacking model provided the most balanced performance across all evaluation metrics and was therefore identified as the most effective classifier for churn prediction in this study.

Table 4

Comprehensive comparison and ranking of classification algorithms

Classification Algorithm	Precision	Accuracy	Recall	F-Measure	Mean Rank	Final Rank
Stacking	86.47	84.07	90.00	86.93	1.75	1
Logistic Regression	85.88	79.90	95.88	87.17	2.75	2
Naïve Bayes	85.59	83.06	89.41	86.12	3.25	3
Support Vector Machine	84.71	82.07	88.82	85.31	4.25	4
Decision Tree	83.24	83.43	82.94	83.19	5.75	5
AdaBoost	84.12	81.87	87.65	84.66	5.75	6
Bagging	72.35	64.62	98.82	78.14	6.25	7
Neural Network	81.18	77.32	88.24	82.42	6.75	8
Nearest Neighbor	66.76	61.83	87.65	72.51	8.50	9
Linear Discriminant Analysis	58.53	57.92	62.35	60.06	10.00	10

Table 4 summarizes the comparative performance of all classification algorithms. The Stacking model achieved the highest precision and accuracy and ranked first overall. Logistic Regression ranked second, largely due to its high recall and F-Measure, while Naïve Bayes ranked third because of its balanced predictive performance. Support Vector Machine ranked fourth, followed by Decision Tree and AdaBoost. Bagging ranked seventh despite its very high recall because its precision and accuracy were relatively low. Neural Network, Nearest Neighbor, and Linear Discriminant Analysis showed weaker performance compared with the other models. The results indicate that the Stacking algorithm is the most suitable model for developing a managerial decision-support system for retail churn prediction.

Association rule mining was then applied to identify product combinations purchased concurrently by important customer segments. Classification and clustering outputs were integrated to identify target customers based on both value and churn status. Their transaction data were reformatted into binary purchase/non-purchase variables. To reduce dimensionality and remove rare products, only items appearing in at least 50% of the target group’s transactions were retained for the final analysis. The FP-Growth algorithm was executed in RapidMiner using Support = 0.5 and Confidence = 0.5 as threshold values.

The key association rule formulas were as follows:

$$Support(A \rightarrow B) = \frac{Count(A \cup B)}{N}$$

$$Confidence(A \rightarrow B) = \frac{Count(A \cup B)}{Count(A)}$$

$$Lift(A \rightarrow B) = \frac{Confidence(A \rightarrow B)}{Support(B)}$$

where Arepresents the premise itemset, Brepresents the conclusion itemset, and Nrepresents the total number of transactions. Support indicates how frequently an itemset occurs in the dataset, Confidence indicates the probability of purchasing Bgiven that $A$was purchased, and Lift indicates the strength of association between $A$and Brelative to random co-occurrence.

Figure 13

Extraction of frequent purchase patterns for target customers using FP-Growth

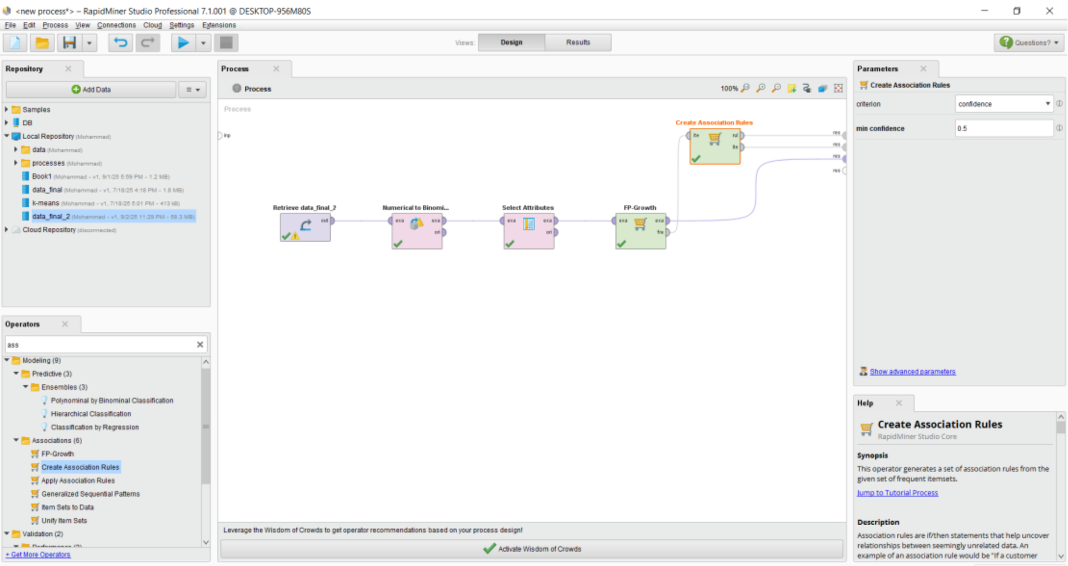


Table 5

Frequent product bundles and co-purchase patterns in loyal customers' shopping baskets

Pattern Type	Size	Support	Item 1	Item 2	Item 3	Item 4
Quadruple bundle	4	0.121	2200170	2203314	2202312	2204359
Quadruple bundle	4	0.127	2200170	2203314	2202312	2204364
Triple co-purchase	3	0.139	2200170	2203314	2202312	-
Triple co-purchase	3	0.121	2200170	2203314	2204359	-
Triple co-purchase	3	0.128	2200170	2203314	2204364	-
Triple co-purchase	3	0.121	2200170	2202312	2204359	-
Triple co-purchase	3	0.127	2200170	2202312	2204364	-
Triple co-purchase	3	0.122	2203314	2202312	2204359	-
Triple co-purchase	3	0.128	2203314	2202312	2204364	-
Product pair	2	0.117	2201035	6260055843036	-	-
Product pair	2	0.140	2200170	2203314	-	-
Product pair	2	0.139	2200170	2202312	-	-
Product pair	2	0.123	2200170	2204359	-	-
Product pair	2	0.128	2200170	2204364	-	-
Product pair	2	0.141	2203314	2202312	-	-
Product pair	2	0.125	2203314	2204359	-	-
Product pair	2	0.130	2203314	2204364	-	-
Product pair	2	0.124	2202312	2204359	-	-
Product pair	2	0.128	2202312	2204364	-	-

Table 5 shows that several product combinations appeared frequently in the shopping baskets of loyal customers. The strongest product-pair support was observed for the combination of 2203314 and 2202312, with support of 0.141, followed by 2200170 and 2203314 with support of 0.140, and 2200170 and 2202312 with support of 0.139. The triple itemset containing 2200170, 2203314, and 2202312 also showed relatively high support of 0.139. These patterns suggest that specific product combinations were repeatedly purchased together by loyal customers and could therefore be used in bundled promotions, personalized offers, and cross-selling strategies.

Top-selling individual items in loyal customers' shopping baskets were also identified. The highest-support individual product was 2201035, with support of 0.332, followed by 6260055843036 with support of 0.295, 6261906101176 with support of 0.220, 2208121 with support of 0.194, 2206861 with support of 0.192, and 6260010500394 with support of 0.190. Other frequent individual products included 6267604414693 (0.186), 164200275300044 (0.176), 2201946 (0.176), 164200275300001 (0.174), 6260043900017 (0.171), 6262916600741 (0.160), 2200134 (0.156), 6260016342011 (0.153), 2200170 (0.152), 2200917 (0.152), 2203314 (0.151), 6265499500040 (0.149), 164200275300002 (0.149), 2201157 (0.149), 6269523900059 (0.147), 2202312 (0.147), 6260459800079 (0.146), 6261847502209 (0.142), 6261061100533 (0.142), 6260053186241 (0.141), 2204359 (0.141), 6260176802929 (0.141), 2201000 (0.138), 6263812801249 (0.138), 164200203250232 (0.138), 2204364 (0.136), 6260105005155 (0.131), 2201655 (0.129), 2200893 (0.129), 6261985200050 (0.126), 6260161529275 (0.126), 6260532822097 (0.125), 2201691 (0.125), 6262157900105 (0.120), 6260667302532 (0.120), 6260053183974 (0.120), 6260707400587 (0.119), 2201014 (0.119), 6263812801386 (0.119), 6260667302525 (0.117), 6260053183998 (0.117), 2201015 (0.117), 6268187102113 (0.116), 6265547620287 (0.115), 2201016 (0.115), and 2201658 (0.115). These individual product supports indicate which items were most central in the consumption patterns of loyal customers.

Figure 14

Visualization of product 2200351 linkages within shopping baskets

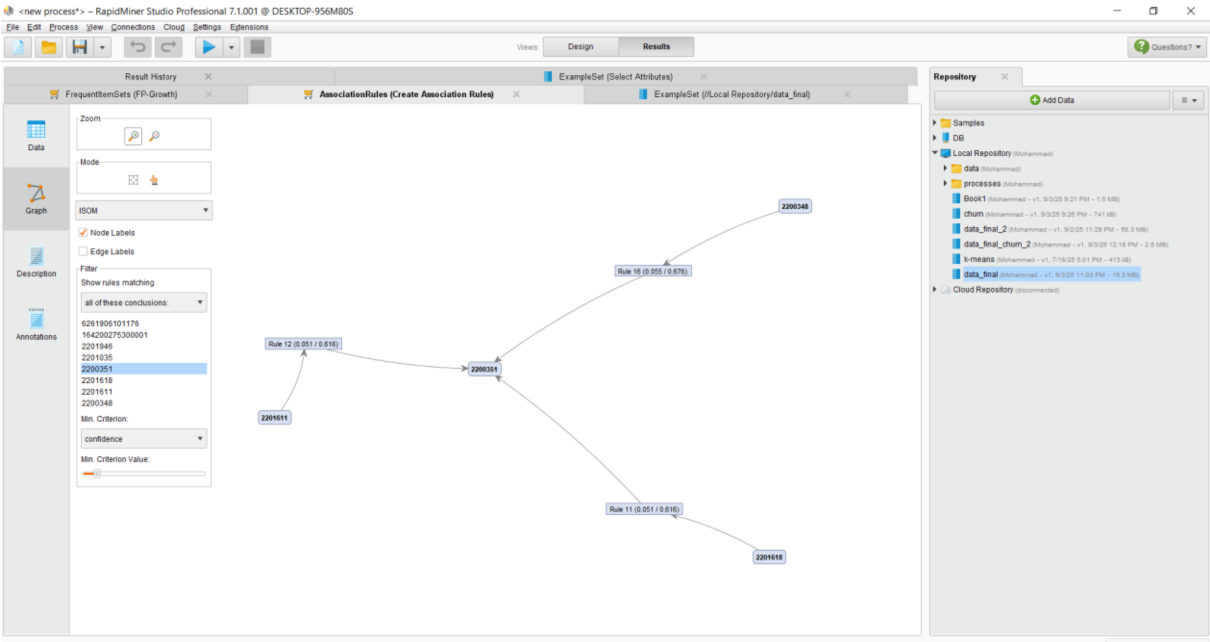


Table 6

Association rules for selected products and high-value churn-risk customers

Segment / Product Focus	Premise	Conclusion	Support	Confidence
Product 2200170	2203314, 2202312	2200170, 2204364	0.13	0.91
Product 2200170	2203314	2200170, 2202312	0.14	0.92
Product 2200170	2203314	2200170	0.14	0.93
Product 2200170	2204364	2200170, 2202312	0.13	0.94
Product 2200170	2204364	2200170, 2203314, 2202312	0.13	0.94
Product 2200170	2204364	2200170	0.13	0.94
Product 2200170	2204364	2200170, 2203314	0.13	0.94
Product 2200170	2202312	2200170, 2203314	0.14	0.94
Product 2200170	2202312	2200170	0.14	0.95
Product 2200170	2203314, 2204359	2200170, 2202312	0.12	0.97
Product 2200170	2203314, 2204359	2200170	0.12	0.97
Product 2200170	2202312, 2204359	2200170, 2203314	0.12	0.98
Product 2200170	2202312, 2204359	2200170	0.12	0.98

Product 2200170	2203314, 2204364	2200170, 2202312	0.13	0.98
Product 2200170	2203314, 2202312, 2204359	2200170	0.12	0.99
Product 2200170	2203314, 2204364	2200170	0.13	0.99
Product 2200170	2203314, 2202312	2200170	0.14	0.99
Product 2200170	2202312, 2204364	2200170	0.13	1.00
Product 2200170	2202312, 2204364	2200170, 2203314	0.13	1.00
Product 2200170	2203314, 2202312, 2204364	2200170	0.13	1.00
High-value churn-risk customers	6261906101176	2201035	-	0.204
High-value churn-risk customers	164200275300001	2201946	-	0.268
High-value churn-risk customers	6261906101176	164200275300001	-	0.328
High-value churn-risk customers	2201946	164200275300001	-	0.338
High-value churn-risk customers	164200275300001	6261906101176	-	0.402
High-value churn-risk customers	2201035	6261906101176	-	0.436
High-value churn-risk customers	2201611	2200348	-	0.562
High-value churn-risk customers	2200348	2201611	-	0.577
High-value churn-risk customers	2200351	2201618	-	0.584
High-value churn-risk customers	2200351	2201611	-	0.584
High-value churn-risk customers	2201618	2200351	-	0.616
High-value churn-risk customers	2201611	2200351	-	0.616
High-value churn-risk customers	2200351	2200348	-	0.623
High-value churn-risk customers	2201618	2201611	-	0.630
High-value churn-risk customers	2201611	2201618	-	0.630
High-value churn-risk customers	2200348	2200351	-	0.676

The association rules in Table 6 show that product 2200170 had strong co-purchase relationships with several other frequently purchased products. The highest confidence values were observed in the rules “2202312, 2204364 → 2200170” and “2203314, 2202312, 2204364 → 2200170,” both with confidence values of 1.00. Other strong rules included “2203314, 2202312 → 2200170” with confidence of 0.99, “2203314, 2204364 → 2200170” with confidence of 0.99, and “2202312, 2204359 → 2200170” with confidence of 0.98. These results indicate that certain products have a highly stable co-occurrence structure in loyal customers’ shopping baskets.

For high-value customers at risk of churn, the strongest rule was “2200348 → 2200351,” with confidence of 0.676. Other notable rules included “2201618 → 2201611” and “2201611 → 2201618,” both with confidence of 0.630, “2200351 → 2200348” with confidence of 0.623, and “2201618 → 2200351” and “2201611 → 2200351,” both with confidence of 0.616. These results show that high-value churn-risk customers continue to show identifiable and repeated product-purchasing patterns. Although product identifiers were anonymized as numeric codes, category-level inspection indicated that many frequent itemsets corresponded to essential and high-turnover product groups such as staple food items, dairy products, and frequently replenished household goods. Therefore, churn in this context does not necessarily indicate disengagement from core consumption needs, but rather reduced interaction frequency, possible switching behavior, or external competitive effects.

Overall, the results show that the proposed RFFML-based framework successfully combined customer segmentation, churn classification, and association rule mining. The clustering stage identified behaviorally meaningful customer groups, the classification stage showed that the Stacking model provided the strongest predictive performance, and the association rule mining stage revealed actionable product-bundle patterns among loyal and churn-risk customers. These findings indicate that the integrated framework can be used as a practical decision-support system for identifying valuable churn-risk customers, designing personalized retention campaigns, creating targeted bundles, and improving customer loyalty in retail stores.

4. Discussion and Conclusion

The findings of the present study demonstrated that sustainable performance management in the Iranian petrochemical industry under sanction conditions is a multidimensional and context-dependent construct that requires simultaneous attention to environmental, economic, governance, social, innovation, and supply chain dimensions. The results obtained from the systematic literature review and fuzzy Delphi analysis indicated that environmental indicators, particularly reduction of greenhouse gas emissions, energy consumption optimization, water recycling, and pollution control, achieved the highest levels of expert consensus. These findings confirm that environmental sustainability is no longer a peripheral concern in petrochemical industries, but rather a strategic operational priority directly associated with efficiency, cost reduction, legitimacy, and long-term resilience. The high prioritization of environmental indicators can be explained by the energy-intensive and emission-intensive nature of petrochemical production processes, where resource inefficiency directly increases operational costs and environmental risks. These findings are aligned with studies emphasizing the importance of integrating environmental sustainability into performance evaluation systems and industrial strategy (Eccles & Klimenko, 2019; Kuhnen & Hahn, 2019; Lu et al., 2018). The findings are also consistent with evidence suggesting that sustainable business performance in petroleum industries depends heavily on environmental management capability and resource optimization (Hadi & Baskaran, 2021).

The strong emphasis placed by experts on energy optimization and eco-efficiency also reflects the economic realities of petrochemical industries operating under sanctions. The findings demonstrated that energy auditing, thermal and electrical integration, reduction of flaring and leakage, heat recovery, and reduction of energy intensity were among the most highly prioritized indicators within the innovative and economical production methods category. This result is understandable because sanctions increase production costs, limit access to advanced technologies, and intensify the need for internal efficiency improvement. Under such conditions, companies are forced to maximize output using existing infrastructure and limited resources. These findings support prior research indicating that eco-efficiency and green lean practices can simultaneously improve environmental and economic performance by reducing waste, enhancing process efficiency, and optimizing resource utilization (Santoso & Kasih, 2024). The results also align with sustainability-oriented performance frameworks emphasizing that operational efficiency and sustainability should be integrated rather than treated as competing objectives (Abbas, 2020). Therefore, environmental sustainability in the Iranian petrochemical industry appears to function not only as an ecological requirement but also as a mechanism for economic survival and operational continuity.

Another important finding of the study was the high importance attributed to digitalization, Industry 4.0 technologies, and intelligent maintenance systems. Indicators related to industrial internet of things, advanced process control, predictive maintenance, digital twins, and integrated production-management systems achieved high defuzzification scores in the fuzzy Delphi stage. This finding indicates that experts perceive digital transformation as one of the principal pathways toward sustainable and competitive petrochemical production under uncertain conditions. The increasing complexity of petrochemical operations requires organizations to improve monitoring capability, predictive decision-making, operational visibility, and process integration. Previous studies have similarly emphasized the transformative role of Industry 4.0 technologies in manufacturing efficiency, operational flexibility, and sustainability performance (Xu et al., 2018). Furthermore, managerial lessons associated with Industry 4.0 implementation demonstrate that organizational readiness, leadership support, infrastructure capability, and strategic integration are essential for successful digital transformation (Sony & Naik, 2019). The present findings also correspond with recent research suggesting that artificial intelligence and smart manufacturing systems can substantially improve sustainability performance through predictive analytics, adaptive optimization, and intelligent resource management (Youssef et al., 2025). In the Iranian petrochemical industry, where sanctions restrict access to imported technologies and increase maintenance difficulties, predictive maintenance and digital process monitoring can reduce downtime, improve operational readiness, and enhance production resilience.

The findings also revealed that process intensification and technological innovation were among the most significant dimensions of innovative and economical production methods. Experts emphasized the importance of compact process technologies, advanced distillation systems, modular design, process integration, and domestic technological alternatives. This result can be interpreted within the broader context of sustainability transitions and industrial adaptation. Petrochemical firms operating under sanctions cannot depend indefinitely on imported technologies, catalysts, and engineering services. Consequently, process redesign and localized technological innovation become essential for maintaining production continuity and improving competitiveness. These findings are consistent with studies highlighting the importance of technological innovation capability in organizational performance and sustainability development (Kim & Jun, 2022). They also correspond with research on sustainability transitions in petroleum-related industries, which indicates that industrial adaptation requires diversification of technological capability and restructuring of innovation systems (Andersen & Gulbrandsen, 2020). In addition, studies on sanctions and green innovation suggest that external restrictions may simultaneously constrain technology access and stimulate domestic innovation capacity (Fu et al., 2023). The findings of the present study strongly support this dual perspective because experts identified localization, substitution of sanctioned materials, and domestically compatible process alternatives as critical indicators for sustainable petrochemical performance.

Supply chain resilience emerged as another highly important dimension in the study. Indicators such as multi-sourcing, domestic supplier development, safety inventory, geopolitical risk management, and business continuity planning received substantial expert agreement. These findings reflect the severe vulnerability of petrochemical supply chains under sanction conditions, where disruptions in procurement channels, logistics systems, or supplier relationships can directly threaten production continuity. The results are consistent with resilience-based supply chain theories emphasizing flexibility, redundancy, adaptability, and risk management as critical components of resilient industrial systems (Ali et al., 2017). The findings also align with simulation-based analyses showing that external disruptions can propagate rapidly across interconnected supply networks and create cascading operational impacts (Ivanov, 2020). In the context of the Iranian petrochemical industry, sanctions create long-term structural uncertainty rather than temporary operational interruptions; therefore, resilience becomes a permanent strategic requirement rather than a short-term emergency response. The importance attributed to domestic supplier development and localization further demonstrates that experts consider local industrial ecosystems essential for sustainable petrochemical operations under restricted international conditions. These findings also correspond with studies emphasizing sustainable supply chain development and prioritization of resilience-related parameters in industrial systems (Shirazi et al., 2024).

The findings additionally showed that governance-related indicators, including transparency, legal compliance, anti-corruption measures, board independence, and cybersecurity management, were highly valued by experts. This result indicates that sustainable performance management in petrochemical industries cannot be achieved solely through technological and operational improvements. Governance quality influences organizational accountability, strategic consistency, stakeholder trust, and risk management capability. Particularly in sanction environments characterized by financial complexity, operational uncertainty, and geopolitical risk, governance mechanisms become critical for maintaining organizational stability and institutional legitimacy. These findings support prior studies emphasizing the importance of climate-related disclosure, governance readiness, and accountability systems in sustainability-oriented corporate management (Eccles & Klimenko, 2019). They also align with sustainability performance measurement research indicating that governance indicators are necessary components of comprehensive sustainability evaluation systems (Kuhnen & Hahn, 2019). Furthermore, the prioritization of industrial cybersecurity reflects the growing digital dependence of petrochemical systems and the increasing exposure of industrial control systems to cyber threats, particularly in politically sensitive and sanction-affected industries.

In the social dimension, occupational health and safety, employee empowerment, training, and workforce participation received relatively high scores, while broader diversity and inclusion indicators achieved comparatively lower priorities. This finding suggests that experts distinguish between operationally critical social indicators and longer-term institutional or cultural dimensions. Occupational safety is particularly important in petrochemical industries due to the hazardous nature of operations, exposure to chemicals, explosion risk, and process safety concerns. Therefore, experts appear to perceive workforce safety and technical capability as directly linked to operational sustainability and productivity. These findings are compatible with studies indicating that human capital and organizational culture significantly influence sustainability performance and operational effectiveness (Santoso & Kasih, 2024). However, the lower prioritization of diversity and inclusion may reflect the immediate pressures of sanction conditions, where organizations prioritize operational continuity and resource management over broader social transformation objectives. Although diversity and inclusion remain important sustainability dimensions in international frameworks, their perceived urgency may decline in contexts characterized by economic restriction and industrial uncertainty.

The results also demonstrated that sustainable innovation and stakeholder accountability were recognized as complementary yet essential dimensions of the proposed framework. Indicators related to green research and development, circular economy practices, sustainability reporting, stakeholder mapping, and responsiveness to complaints achieved acceptable consensus levels among experts. These findings reinforce the idea that sustainable petrochemical performance requires continuous innovation and active stakeholder engagement rather than isolated operational interventions. Integrated thinking approaches similarly emphasize that sustainability should be embedded across organizational strategy, reporting systems, and value creation processes (Adams, 2017). Furthermore, the growing importance of investor-oriented sustainability expectations and environmental accountability suggests that firms increasingly need credible sustainability indicators to maintain legitimacy and strategic adaptability (Eccles & Krzus, 2019). In the Iranian petrochemical context, stakeholder accountability may also strengthen social trust, institutional credibility, and long-term competitiveness despite external restrictions.

Overall, the findings of the present study indicate that sustainable performance management in the Iranian petrochemical industry under sanctions is fundamentally shaped by the interaction between sustainability imperatives, technological adaptation, economic constraints, and resilience requirements. Unlike generic sustainability models developed for stable and highly globalized industrial contexts, the proposed framework reflects the operational realities of a sanction-affected industrial environment. The identified indicators combine global sustainability principles with localized priorities such as sanction resilience, domestic supplier development, process intensification, substitution of restricted materials, and operational flexibility. Therefore, the study contributes theoretically by contextualizing sustainable performance management within constrained industrial systems and contributes practically by providing a localized framework for petrochemical decision-makers. The findings also demonstrate that sustainable performance and innovative production methods are mutually reinforcing dimensions rather than separate managerial domains. Companies that improve energy efficiency, digital capability, process integration, innovation systems, and supply chain resilience are simultaneously strengthening sustainability performance, operational continuity, and long-term competitiveness.

One limitation of the present study is that the fuzzy Delphi panel was limited to 18 experts from the petrochemical industry and academia, which may restrict the generalizability of the findings to all segments of the energy and process industries. In addition, the study adopted a qualitative-exploratory design and did not empirically test the causal relationships among the identified indicators. Another limitation is that the prioritization of indicators may change over time depending on geopolitical developments, technological evolution, regulatory changes, and the intensity of international sanctions. Furthermore, because the study focused specifically on the Iranian petrochemical industry, the identified dimensions and priorities may not fully apply to industries with different technological structures or institutional environments.

Future research can expand the proposed framework through quantitative validation methods such as structural equation modeling, interpretive structural modeling, DEMATEL, or system dynamics approaches. Researchers may also investigate the causal interactions among sustainability dimensions, digitalization capability, supply chain resilience, and innovation performance. Comparative studies between sanction-affected and non-sanctioned petrochemical industries could provide additional insights into the contextual nature of sustainable performance management. Future studies may further explore the role of artificial intelligence, smart manufacturing systems, circular economy models, and climate-related governance in transforming petrochemical sustainability practices. Longitudinal studies are also recommended to examine how sustainability priorities evolve over time under changing technological and geopolitical conditions.

From a practical perspective, petrochemical managers and policymakers should prioritize the integration of sustainability indicators into operational and strategic decision-making systems. Companies should invest in energy optimization, predictive maintenance, digital infrastructure, and process intensification to improve operational efficiency and resilience. Policymakers should support domestic supplier development, technology localization, and industrial innovation ecosystems to reduce dependence on restricted foreign technologies. Petrochemical firms should also strengthen governance systems, transparency mechanisms, industrial cybersecurity, and sustainability reporting practices in order to improve institutional trust and long-term competitiveness. Furthermore, organizations should establish integrated sustainability management systems that combine environmental responsibility, operational efficiency, innovation capability, and supply chain resilience within a unified strategic framework.

Authors' Contributions

Authors contributed equally to this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

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

Ethics Considerations

In this research, ethical standards including obtaining informed consent, ensuring privacy and confidentiality were considered.

References

Abbas, J. (2020). Impact of Total Quality Management on Corporate Green Performance through the Mediating Role of Corporate Social Responsibility. *Journal of Cleaner Production*, *242*, 118458. https://doi.org/10.1016/j.jclepro.2019.118458

Abdi, F., & Abolmakarem, S. (2019). Customer Behavior Mining Framework (CBMF) Using Clustering and Classification Techniques. *Journal of Industrial Engineering International*, *15*(Suppl 1), 1-18. https://doi.org/10.1007/s40092-018-0285-3

Adams, C. (2017). *The Sustainable Development Goals, Integrated Thinking and the Integrated Report*.

Agbemadon, K. B., Couturier, R., & Laiymani, D. (2022). Churn Detection Using Machine Learning in the Retail Industry. International Conference on Computer, Control and Robotics (ICCCR),

Ali, A., Mahfouz, A., & Arisha, A. (2017). Analysing Supply Chain Resilience: Integrating the Constructs in a Concept Mapping Framework via a Systematic Literature Review. *Supply Chain Management: An International Journal*, *22*(1), 16-39. https://doi.org/10.1108/SCM-06-2016-0197

Andersen, A. D., & Gulbrandsen, M. (2020). The Innovation and Industry Dynamics of Technology Phase-Out in Sustainability Transitions: Insights from Diversifying Petroleum Technology Suppliers in Norway. *Energy Research & Social Science*, *64*, 101447. https://doi.org/10.1016/j.erss.2020.101447

Bagul, N., Berad, P., Surana, P., & Khachane, C. (2021). Retail Customer Churn Analysis Using RFM Model and K-Means Clustering. *International Journal of Engineering Research and Technology*, *10*(03).

Bhargava, D., Bhargava, A., Melgarejo-Bolivar, R. P., Montes de Oca-Nina, A. M., & Chaudhury, S. (2024). Enhancing Personalization of Customer Services in E-Commerce System Using Predictive Analytics. *Fluctuation and Noise Letters*, *23*(02), 2440011. https://doi.org/10.1142/S021947752440011X

Casteran, H., Meyer-Waarden, L., & Reinartz, W. (2017). Modeling Customer Lifetime Value, Retention, and Churn. In *Handbook of market research* (pp. 1-33). Springer. https://doi.org/10.1007/978-3-319-05542-8_21-1

Chen, J., Guo, Z., Zhang, C., Tian, Y., & Li, Y. (2023). Research on the Application of CEEMD-LSTM-LSSVM Coupled Model in Regional Precipitation Prediction. *Water*, *15*(8), 1465. https://doi.org/10.3390/w15081465

Cheng, Z., He, J., Xu, S., & Yang, X. (2023). Coupling Assessment for the Water-Economy-Ecology Nexus in Western China. *Ecological Indicators*, *154*, 110648. https://doi.org/10.1016/j.ecolind.2023.110648

Eccles, R. G., & Klimenko, S. (2019). The Investor Revolution. *Harvard business review*, *97*(3), 106-116.

Eccles, R. G., & Krzus, M. P. (2019). Implementing the Task Force on Climate-Related Financial Disclosures Recommendations: An Assessment of Corporate Readiness. *Schmalenbach Business Review*, *71*(2), 287-293. https://doi.org/10.1007/s41464-018-0060-4

Fu, Q., Gong, Q., Zhao, X. X., & Chang, C. P. (2023). The Effects of International Sanctions on Green Innovations. *Technological and Economic Development of Economy*, *29*(1), 141-164. https://doi.org/10.3846/tede.2022.17782

Gaur, A., & Dubey, R. (2018). Predicting Customer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques. 2018 International Conference on Advanced Computation and Telecommunication (ICACAT),

Hadi, S., & Baskaran, S. (2021). Examining Sustainable Business Performance Determinants in Malaysia Upstream Petroleum Industry. *Journal of Cleaner Production*, *294*, 126231. https://doi.org/10.1016/j.jclepro.2021.126231

Hughes, A. M. (1994). *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable, Customer-Based Marketing Program*.

Ivanov, D. (2020). Predicting the Impacts of Epidemic Outbreaks on Global Supply Chains: A Simulation-Based Analysis on the Coronavirus Outbreak (COVID-19/SARS-CoV-2) Case. *Transportation Research Part E: Logistics and Transportation Review*, *136*, 101922. https://doi.org/10.1016/j.tre.2020.101922

Khandelwal, N., & Sakalle, D. V. (2025). TBMOR Customer Churn Predication Using XGBoost Classifier. *Journal of Information Systems Engineering and Management*. https://doi.org/10.52783/jisem.v10i19s.3111

Kim, Y., & Jun, H. (2022). Exploring Technology Innovation Factors, Government Support and Performance of Development-Related Social Enterprises: Evidence from South Korea. *Sustainability*, *14*(22), 15406. https://doi.org/10.3390/su142215406

Kuhnen, M., & Hahn, R. (2019). From SLCA to Positive Sustainability Performance Measurement: A Two-Tier Delphi Study. *Journal of Industrial Ecology*, *23*(3), 615-634. https://doi.org/10.1111/jiec.12762

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer Churn Prediction System: A Machine Learning Approach. *Computing*, *104*(2), 271-294. https://doi.org/10.1007/s00607-021-00908-y

Lu, M. T., Hsu, C. C., Liou, J. J., & Lo, H. W. (2018). A Hybrid MCDM and Sustainability-Balanced Scorecard Model to Establish Sustainable Performance Evaluation for International Airports. *Journal of Air Transport Management*, *71*, 9-19. https://doi.org/10.1016/j.jairtraman.2018.05.008

Malhotra, R., Shah, S., & Bansal, S. (2025). Customer Churn Analysis Using Machine Learning in Telecom Industry. Proceedings of Data Analytics and Management, Singapore.

Matuszelanski, K., & Kopczewska, K. (2022). Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(1), 165-198. https://doi.org/10.3390/jtaer17010009

Mendez, G. A. D. (2020). *Data Mining Guided Process for Churn Prediction in Retail: From Descriptive to Predictive Analytics* Universidade NOVA de Lisboa]. Portugal.

Mining, W. I. D. (2006). *Data Mining: Concepts and Techniques* (Vol. 10). Morgan Kaufmann.

Pasquadibisceglie, V., Appice, A., Ieva, G., & Malerba, D. (2024). TSUNAMI: An Explainable PPM Approach for Customer Churn Prediction in Evolving Retail Data Environments. *Journal of Intelligent Information Systems*, *62*(3), 705-733. https://doi.org/10.1007/s10844-023-00838-5

Prabadevi, B., Shalini, R., & Kavitha, B. R. (2023). Customer Churning Analysis Using Machine Learning Algorithms. *International Journal of Intelligent Networks*, *4*, 145-154. https://doi.org/10.1016/j.ijin.2023.05.005

Rahim, M. A., Mushafiq, M., Khan, S., & Arain, Z. A. (2021). RFM-Based Repurchase Behavior for Customer Classification and Segmentation. *Journal of Retailing and Consumer Services*, *61*, 102566. https://doi.org/10.1016/j.jretconser.2021.102566

Santoso, B., & Kasih, T. P. (2024). Green Lean Concept for Measurement of Sustainable Performance Mediated by Organizational Culture in Oil & Gas and Petrochemical Industry. *International Review of Management and Marketing*, *14*(5), 88. https://doi.org/10.32479/irmm.16621

Schaeffer, S. E., & Sanchez, S. V. R. (2020). Forecasting Client Retention: A Machine-Learning Approach. *Journal of Retailing and Consumer Services*, *52*, 101918. https://doi.org/10.1016/j.jretconser.2019.101918

Shirazi, A., Mohammadi Dolat-Abadi, H., Dehghani, J., & Asgari Sooran, M. (2024). Prioritizing Key Parameters for Sustainable Supply Chain Development in the Construction Industry Using a Copeland Hybrid Method. *Environmental Energy and Economic Research*, *8*(4), eS095.

Sony, M., & Naik, S. S. (2019). Ten Lessons for Managers While Implementing Industry 4.0. *IEEE Engineering Management Review*, *47*(2), 45-52. https://doi.org/10.1109/EMR.2019.2913930

Valente, F., Paredes, S., Henriques, J., Rocha, T., de Carvalho, P., & Morais, J. (2022). Interpretability, Personalization and Reliability of a Machine Learning Based Clinical Decision Support System. *Data Mining and Knowledge Discovery*, *36*(3), 1140-1173. https://doi.org/10.1007/s10618-022-00821-8

Vemulapalli, G. (2024). AI-Driven Predictive Models Strategies to Reduce Customer Churn. *International Numeric Journal of Machine Learning and Robots*, *8*(8), 1-13.

Verma, A. (2020). Consumer Behaviour in Retail: Next Logical Purchase Using Deep Neural Network. *ArXiv*. https://arxiv.org/abs/2010.06952

Wu, P., Wang, Z., Li, H., & Zeng, N. (2024). KD-PAR: A Knowledge Distillation-Based Pedestrian Attribute Recognition Model with Multi-Label Mixed Feature Learning Network. *Expert Systems with Applications*, *237*, 121305. https://doi.org/10.1016/j.eswa.2023.121305

Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: State of the Art and Future Trends. *International Journal of Production Research*, *56*(8), 2941-2962. https://doi.org/10.1080/00207543.2018.1444806

Youssef, A. B., Omrani, N., & Zeqiri, A. (2025). AI for Good: Advancing Sustainability in Smart Manufacturing. *Ieee Transactions on Engineering Management*. https://doi.org/10.1109/TEM.2025.3640227