Future of Work and Digital Management Journal

Article type:
Original Research

Article history:
Received 11 February 2025
Revised 21 May 2025
Accepted 26 May 2025
Published online 01 June 2025

Hybrid k-means and SVM machine learning for B2B customer segmentation:A case study banking for sustianable sales

Seyyed Amir Masoud. Homayooni [1], Mahdi. Mohamadi [2*]

1 Department of Digital Electronic Systems, University of Science and Technology, Tehran, Iran
2 Department of Economics, Imam Sadiq University, Tehran, Iran

Corresponding author email address:
mahdimohamadi5060@gmail.com

How to cite this article:
Homayooni, S. A. M., & Mohamadi, M. (2025). Hybrid k-means and SVM machine learning for B2B customer segmentation:A case study banking for sustianable sales . *Future of Work and Digital Management Journal*, *3*(2), 1-17.
https://doi.org/10.61838/10.61838/fwdmj.3.2.6

ABSTRACT

The effectiveness of segmenting Business-to-Business (B2B) customers is necessary to sales strategies and optimize resources. Several clustering methods have been documented in the literature, although limited research has studied the application of machine learning methods for sustainable sales optimization. This study addresses this gap by applying K-means and X-means clustering and Support Vector Machine (SVM) classification, to analyze and segment Bank Mellat B2B customers. Two data analysis methods were used as measures of the customers' transactional and behavioral activities, which included purchase history, frequency of interactions, and service usage. The analysis resulted in two segments of customer identified in the dataset: Cluster 0 consists of customers characterized by low engagement who are younger, and customers with less financial activity; Cluster 1 consists of customers characterized by high levels of engagement and are older while having high account balance and loyalty. The clustering method had an 89% degree of accuracy and SVM had 90% degree of accuracy based the final clustering and was validated with sensitivity analysis. These measures provide improved analytics and enable targeted engagement for further sustainability engagement, such as customer digital engagement for Cluster 0, Cluster 1 with loyalty programs are two targeted engagement methods enhancing resource optimization and further engagement by reducing physically-based interactions. This hybrid approach ultimately provided better engagement analytics measuring both unsupervised and supervised learning analyzed within the static and dynamic outcomes of the bank's data while offering a better scalable solution for profitability after customer engagement and retention. Future research imports an additional measure of sustainability inquiry examining tonsuring sustainability metrics and impacts through machine learning (ML) modeling.

Keywords: B2B customer analysis, clustering, machine learning, sustainable sales, resource allocation, sales strategies, machine learning algorithms, customer loyalty, resource optimization, customer groups

Introduction

In today's highly competitive and data-driven economy, business-to-business (B2B) organizations are increasingly compelled to leverage advanced analytical tools to understand and serve their customers more effectively. Unlike business-to-consumer (B2C) transactions, which are often driven by emotional and impulse-based decisions, B2B exchanges are characterized by extended decision-making cycles, relational complexities, and higher stakes in terms of value delivery and long-term partnership outcomes [1]. For this reason, customer segmentation—the process of dividing customers into homogeneous groups based on shared characteristics—plays a crucial role in shaping marketing strategies, resource allocation, and sustainable sales practices [2]. By enabling firms to target the right customer segments with tailored solutions, segmentation strengthens competitiveness and fosters more resilient business models.

The concept of sustainability in sales has further amplified the importance of segmentation. Organizations are now challenged not only to maximize short-term revenues but also to secure long-term relational capital, customer loyalty, and efficient resource utilization [3]. The Triple Bottom Line (TBL) perspective emphasizes that businesses must balance economic, environmental, and social considerations in their strategies [4]. In B2B contexts, sustainable segmentation provides a pathway for companies to align with the TBL while maintaining profitability and market responsiveness [5]. The banking sector, in particular, faces heightened expectations regarding customer-centric sustainability strategies, where segmentation can optimize service offerings, reduce wasted effort, and ensure better targeting for diverse client bases.

Machine learning (ML) has emerged as a pivotal enabler of this transformation, providing advanced techniques for pattern recognition, predictive analytics, and automated decision-making [6]. Compared with traditional statistical approaches, ML allows for the analysis of large and complex datasets, uncovering latent behavioral patterns and producing actionable insights with higher accuracy [7]. In the domain of customer segmentation, ML clustering algorithms such as k-means and fuzzy C-means have proven particularly effective at identifying hidden structures in customer data [8, 9]. These methods can be combined with classification models like Support Vector Machines (SVM) to enhance predictive capabilities, enabling firms to not only cluster customers based on past behavior but also predict future engagement levels [10].

The literature reflects an expanding body of research on ML-based customer segmentation. Studies have demonstrated the ability of clustering algorithms to produce more accurate and granular segmentation compared to conventional techniques like Recency, Frequency, and Monetary (RFM) analysis [11]. For example, Hadid et al. proposed a hybrid method combining RFM with graph models, incorporating firmographic data to improve segmentation relevance in B2B contexts [12]. Similarly, Li et al. advanced the traditional RFM approach by applying weighted k-means clustering in e-commerce, achieving more precise grouping of customers [13]. These studies illustrate the necessity of expanding beyond simple transactional measures to include multidimensional data sources in segmentation frameworks.

The role of data-driven modeling in understanding customer behavior has been extensively discussed, with scholars highlighting its potential to enhance both customer relationship management (CRM) and strategic marketing decisions [14]. Data-driven decision-making frameworks emphasize that customer analytics must move beyond descriptive assessments to predictive and prescriptive functions, aligning with sustainable business objectives [15]. Within this scope, researchers have also highlighted the adoption challenges of ML-based analytical tools in digital marketing, stressing organizational readiness and managerial support as critical for success [6]. These observations highlight that while technology adoption is growing, a gap remains in systematically integrating ML segmentation into long-term strategic sustainability practices.

The development of machine learning-based segmentation has also been linked to broader theoretical perspectives in marketing and strategy. The Resource-Based View (RBV) suggests that unique analytical capabilities, such as advanced data mining and ML algorithms, can serve as strategic resources that create sustainable competitive advantage [16]. By leveraging firm-specific data and applying sophisticated analytic methods, companies can generate valuable, rare, and inimitable insights that strengthen market positioning. Meanwhile, Relationship Marketing Theory underscores the importance of commitment, trust, and mutual value creation in sustainable customer engagement [4]. Combining RBV and Relationship Marketing perspectives provides a theoretical justification for embedding ML-driven segmentation as a key organizational capability for sustaining long-term sales growth.

Several empirical studies demonstrate the effectiveness of ML in customer segmentation and predictive analytics. Wisesa et al. applied ML techniques to predict B2B sales of telecommunication services, showing significant accuracy improvements over conventional methods [17]. Similarly, Purnomo et al. used clustering techniques to derive more effective marketing strategies by identifying high-value customers in B2B settings [18]. Research by Abidar et al. highlighted the potential of customer segmentation with ML for designing targeted marketing actions, underscoring the practical utility of such methods in improving ROI [19]. These studies collectively affirm the growing role of ML in transforming segmentation from a descriptive exercise into a predictive and prescriptive strategic tool.

The methodological landscape has further evolved with comparative analyses of clustering algorithms. Sakina et al. conducted a systematic evaluation of k-means, hierarchical clustering, fuzzy C-means, and DBSCAN, concluding that algorithm selection substantially influences segmentation outcomes and business applicability [20]. Complementing this, Agrawal and Agarwal compared deep learning and machine learning algorithms in medical applications, demonstrating the relative strengths of each in terms of precision and computational efficiency [21]. While outside direct marketing, such comparative studies underscore the importance of methodological rigor in selecting the right algorithms for segmentation tasks.

In addition, domain-specific research has enriched the conversation on segmentation and sustainability. Manzoor et al. reviewed ML methods for customer churn prediction, offering recommendations for business practitioners to apply predictive analytics in retaining valuable customers [22]. Han et al. investigated how livestream studio environments influence sales performance through ML-based analysis, highlighting the broader role of AI-driven insights in shaping customer engagement strategies [23]. Horng and Yenradee extended this perspective to delivery service management in SMEs, demonstrating how ML-supported decision-making can improve service efficiency [24]. Together, these studies suggest that segmentation should be viewed as part of a larger ecosystem of AI-driven tools that contribute to business sustainability.

Despite these advances, several research gaps remain. Many studies have focused on short-term sales optimization or marketing effectiveness, without adequately addressing the long-term sustainability implications of segmentation [25]. Moreover, while clustering methods such as k-means and fuzzy C-means have been widely studied, their integration with supervised models like SVM for predictive classification remains underexplored in the B2B context [26]. Few studies explicitly embed sustainability metrics—such as resource efficiency, digital engagement, or environmental impact—within segmentation frameworks [27]. There is also limited empirical evidence from banking contexts, where customer heterogeneity, regulatory considerations, and sustainability mandates intersect in unique ways [28].

Against this backdrop, the present study proposes a hybrid approach that combines k-means clustering with SVM classification to segment B2B customers in the banking sector, specifically focusing on sustainable sales outcomes. By applying these methods to transactional and behavioral data from Bank Mellat, the research identifies distinct customer segments characterized by varying levels of engagement and loyalty. This dual methodological framework—integrating unsupervised clustering with supervised classification—provides both descriptive and predictive insights, enhancing the capacity of banks to allocate resources strategically and design sustainability-oriented engagement strategies. Importantly, the study situates segmentation within the broader framework of sustainable business practices, thereby bridging the gap between advanced analytics and long-term organizational competitiveness [29-33].

In doing so, this research makes three primary contributions. First, it advances methodological innovation by demonstrating the combined utility of clustering and classification in B2B segmentation. Second, it incorporates sustainability metrics into the segmentation framework, responding to calls for greater integration of environmental and social considerations into marketing analytics. Finally, it provides context-specific insights for the banking industry, where sustainable sales strategies are critical for balancing profitability, compliance, and social responsibility. By addressing these gaps, the study contributes both theoretically and practically to the evolving discourse on ML-driven segmentation for sustainable sales.

Methods and Materials

Based on the dataset, business-to-business (B2B) customers are categorized by real and legal entities. Most of the variables we obtained from the various internal systems of Bank Mellat contained transactional, service usage, as well as some characteristics of customer bases (purchase frequency, interactions in history, etc.). The point to focus is the segmentation of consumers into different clusters depending on their behavior patterns, with the utility purpose of a better lifetime of relationship management in mind.

In this research, machine-learning techniques were used to address the construction of B2B customers in Bank Mellat for the optimization of resources allocation toward sustainable sales strategies. The clustering algorithm applied is the K-means, which is executed through RapidMiner software. to segment the customers into specific group based on a range of variables, such as transaction frequency, the type of service delivered, and their interaction history (i.e., number of marketing calls, days since last contact). The X-means operator was included in the analysis to determine the number of clusters accuracy and objectively and based on a statistical criterion, including the Bayesian Information Criterion (BIC)This process consisted of: (1) data preprocessing to ensure data quality and consistency, (2) first clustering customers into groups using K-means, and (3) the finalization of clusters using the X-means method to arrive at the maximum number of clusters.

The X-means operator was employed to maximize clustering accuracy while providing an automated method for determining the best number of clusters following statistical criteria, like the Bayesian Information Criterion (BIC).1. Data Collection and Initial Description This dataset consists of transactional and behavioral histories of 45,211 business-to-business (B2B) customers of the Bank Mellat.

1.The variables included: Demographic characteristics: Age, type of work, marital status, educational attainment. Behavioral/transactional variables: Account balance, duration of phone call, number of days since last contact (pdays), number of previous interactions, marketing campaign contacts, and service usage history.

target variable: Engagement level of customer (binary: high or low interaction).2. Data Cleaning. Missing Value Handling

2. Data Cleaning

a. Missing Value Handling

Missing values in categorical features (i.e., job type, education level) were imputed with the mode of each feature.

To reduce the effect of outliers, missing numeric data (i.e., account balance and call duration) were imputed with the median.

Records containing >30% missing data (0.8% of the dataset) were removed to maintain data integrity

b. Outlier Detection and Treatment

Interquartile Range (IQR): Employed for numerical features (e.g., account balance, call duration). Values exceeding 1.5×IQR were cut down to the 5th and 95th percentiles.

Numerical Feature Standardization: Z-score normalization was used to standardize numerical features (age, account balance, call duration, etc.) to have the same scaling.

The formula for standardization is given by:

$$\acute{X} = \frac{X - \mu}{\delta}$$

where $\acute{X}$ is the normalized value and X is the original value; μ is the mean, and δ is the standard deviation of the feature.

c. Duplicate Removal

Duplicate records (1.2% of the records) were identified and removed using unique customer IDs.

3.future selection::Domain Relevance: The variables pdays, campaign, and account balance were kept due to their conceptual relevance to B2B engagement, as justified by prior research that associates these variables with customer loyalty.

Correlation Analysis: The variables having Pearson correlation |r|>0.85|r|>0.85 (e.g., redundant use of services measures) were eliminated to avoid multicollinearity.

4.data transformation

a.categorical variable encoding One-Hot Encoding:

Used on nominal variables (i.e., job type, marital status, contact method). For instance

Job type (6 categories) → 6 binary columns.

Marital status (3 categories) → 3 binary columns.

High-cardinality features (e.g., level of education) were aggregated into more general categories (e.g., "Basic," "Intermediate," "Advanced") prior to encoding.

5 Target Variable Conversion for SVM Classification

In order to enable supervised learning with Support Vector Machines (SVM), the target variable (customer engagement level) was transformed into binary classification:

Engaged Customers (High Interaction) → Label: 1

Less Engaged Customers (Low Interaction) → Label: 0

6. Data Splitting

The dataset was divided into 80% training (36,169 samples) and 20% testing (9,042 samples) sets using stratified sampling to preserve class distribution.

7. Class Imbalance Mitigation

The target variable was imbalanced (70% low-interaction and 30% high-interaction).

SMOTE (Synthetic Minority Oversampling Technique) was utilized in the training set to handle class imbalance (70% low-interaction and 30% high-interaction). Based on the use of k=5 nearest neighbors, SMOTE created synthetic samples for the minority class (high-interaction) to make its representation equivalent to that of the majority class to enhance model performance.

8. Exploratory Data Analysis (EDA)

Visualizations, namely centroid plots and heatmaps (Figures 4–6), were used to identify trends in the distribution of features among various clusters.

Statistical metrics, i.e., mean, variance, and skewness, were calculated for important variables (e.g., balance, age) to guide preprocessing choices.

Implementation Tools

RapidMiner Studio 9.10: Used for K-means/X-means clustering, one-hot encoding, and Z-score normalization.

Python (Scikit-learn): Implemented for SVM, SMOTE, and train-test splitting.

Model validation

The model validation was conducted thrsensitivity analysis to test the impact of data volatility on the results of clustering, along with a performance check using metrics such as accuracy, precision, and recall. The results were cross-checked with actual data to ensure their applicability in the real world. The empirical method provides actionable advice for tailoring Bank Mellat's marketing strategies, optimizing the allocation of resources, and cultivating long-term customer relationships within the B2B sector. The approach serves as a replicable model for similar organizations within similar markets.

SVM method

This study explores the preprocessing steps and performance of a Support Vector Machine (SVM) model in classifying customer interaction levels with a bank.

Support Vector Machine (SVM) Classification Model

The Support Vector Machine (SVM) is a robust supervised learning algorithm for classification, ideal for high-dimensional data. It constructs an optimal hyperplane to separate classes, maximizing the margin. The optimization problem is defined as:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum \xi_i$$

subject to:

$$y_i(w.x_i + b) \geq 1 - \xi_i \, , \xi_i \geq 0$$

where w is the weight vector, b is the bias, C (set to 1.0 in this work) is the regularization parameter, and ξ_i are slack variables. In the case of non-linearly separable data, we used a radial basis function (RBF) kernel:

$$K\left(x_i, x_j\right) = \exp(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2})$$

with σ tuned to 0.1 via grid search to balance model complexity and accuracy.

Evaluation Metrics

For the performance analysis of each model, several statistical metrics were employed. Accuracy is the ratio of correct predictions to all instances. Sensitivity (Recall) is the ability of the model to correctly identify true positive instances, and Precision is the proportion of true positives to all positive predictions. The F1 Score gives the harmonic average of precision and recall and is thus appropriate for imbalanced data. In addition, the Confusion Matrix was also utilized in plotting the model's classification performance in relation to the true positive, false positive, true negative, and false negative values. It provides a detailed overview of classification results, enabling additional information on model behavior.

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

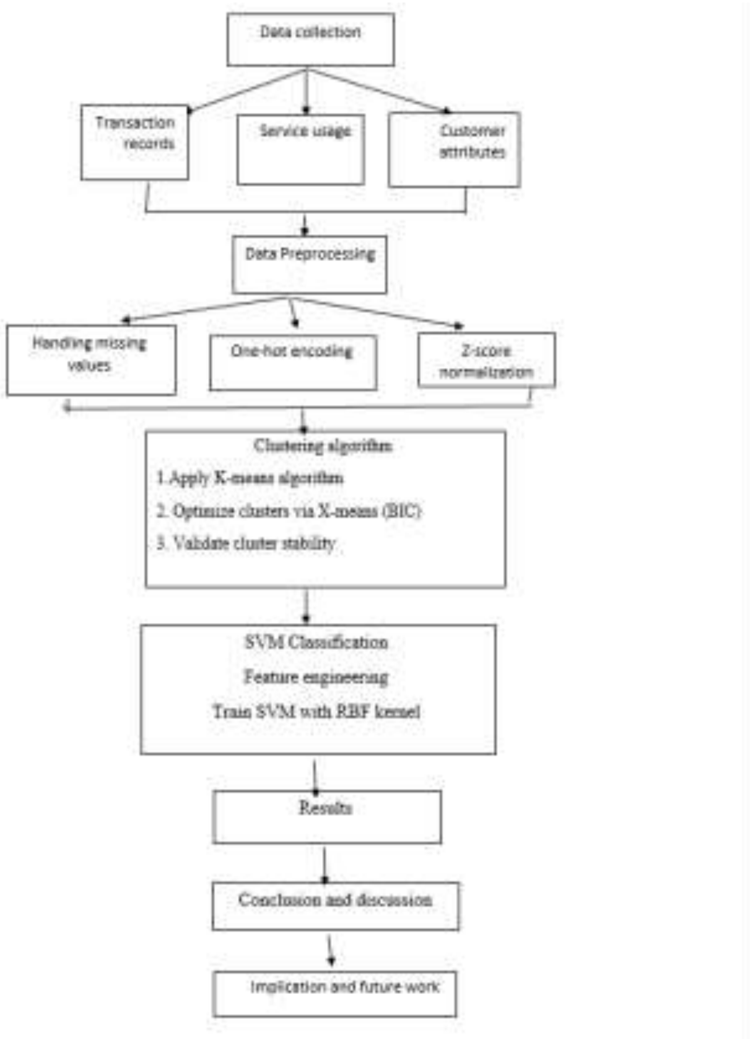
$$Sensitivity\ (Recall) = \frac{TP}{TP + FN}$$

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

$$F1\ Score = 2 \times \frac{Precision\ .\ Recall}{Precision + Recall}$$

Figure 1.

Conceptual model



Findings and Results

The results arise from the application of Machine Learning techniques including K-means clustering algorithm applied to Bank Mellat's B2B customers. The intention was to cluster customers considering their behavioral activities, such as usage patterns, history of engagement, and behaviors relating to purchase activities. This way, the bank is guaranteed to allocate their resources correctly and create sales methodologies for each segment of their customers. The sections below will delve further into the detail and describe the results

Customer Clustering Results The K-means clustering algorithm revealed a behavioral and interactional clustering of the Bank Mellat customers and categorized them into two main clusters the distribution of each cluster is shown in the table below:

• Cluster 0: This group has a larger number of customers with an estimate of 27,167 clients.

• Cluster 1: This group consisted of 18,044 total customers. They appear to be a smaller group of customers, but more responsive than the prior cluster. Both clusters were built from the following variables:

• Pdays: Variation that indicated the number of days since the last contact with the bank. • previous: A variable that showed the count of the interactions they had in the past.

• campaign: The total number of times the marketing department called customers to attempt to win back their business

Characteristics of Each Cluster

Characteristics of Each ClusterThe implications resulting from the study did discern particular characteristics for each cluster that are crucial for a proper understanding of customer attitudes and thoughtfulness.

Cluster 0: Customers with Low Engagement• Behavior: The bank has had relatively little engagement with customers from this cluster. The average number of days since the last contact (pdays) is much higher, which means there tends to be less customer communication from the bank.

• Marketing: The bank has initiated fewer marketing calls (campaign) directed toward this group, which signals that the group is not being optimally targeted.

• Customer Group Profile: Customers in this group are generally younger (36.5 years old on average) and have a relatively low account balance (903,593 currency units on average). For this reason, the customers in this group are new customers or do not have high engagement with the bank.• Interaction: The average length of calls with these customers is lower (278.235 seconds), indicating that more customers in the bank had less meaningful interactions.

- Analysis with a Heatmap

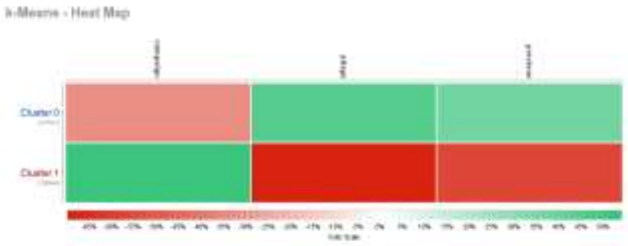
The distinct clusters were also examined using Heatmap visualization. The Heatmap depicts customer distinctions in behavior to interact with the firm or company:

• cluster0 (Low-interaction Customers): The Heatmap shows a dark shade regarding pdays meaning the last time a contact was made to contact the customer was a long time ago while the campaign variable is brighter which shows less contact by the marketing department.

• cluster1 (High-interaction Customers): The Heatmap shows a lighter shade with respect to pdays meaning there was contact made with the customer not long, whereas the campaign variable is darker, which means that the marketing department made a high amount of contacts with the customer.

Figure 2.

K-Means heat map



The Centroid Charts (refer to Figure 5 and Figure 6) present the mean values for important variables in each cluster visually, which serve to provide a straightforward comparison of Cluster 0 and Cluster 1. These charts are relevant to the understanding of customer differences and their impact on important decisions.

Table 1.

Description of Aspects

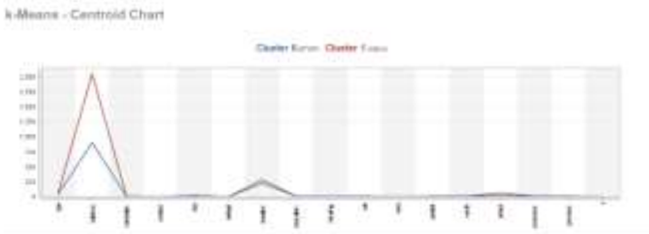
Aspect	Cluster 0: Low-interaction Customers	Cluster 1: High-interaction Customers
Demographics	- Average Age: 36.5 years (younger) - Average Account Balance: 903,593 currency units (lower)	- Average Age: 47.6 years (older) - Average Account Balance: 2,052,986 currency units (higher)
Behavioral Patterns	- Days since last contact (pdays): Higher (less frequent interaction) - Previous contacts: Higher (more past efforts by the bank) - Marketing calls (campaign): Lower (less effort to attract) - Call Duration: Shorter (278.235 seconds)	- Days since last contact (pdays): Lower (more frequent interaction) - Previous contacts: Lower (fewer past efforts by the bank) - Marketing calls (campaign): Higher (more effort to attract/retain) - Call Duration: Longer (227.937 seconds)
Analysis	- Less interaction with the bank - Likely newer or less active customers - Requires targeted marketing and digital services to increase engagement	- More frequent and deeper interactions with the bank - Likely loyal customers using complex services (e.g., loans) - Requires personalized services and loyalty programs
Strategic Focus	- Increase interaction through personalized offers and digital tools - Provide financial advice to younger customers	- Strengthen loyalty through reward programs - Offer specialized financial products and services

Figure 3.

Centroid Chart for X-Means

Figure 4.

Centroid Chart for X-Means



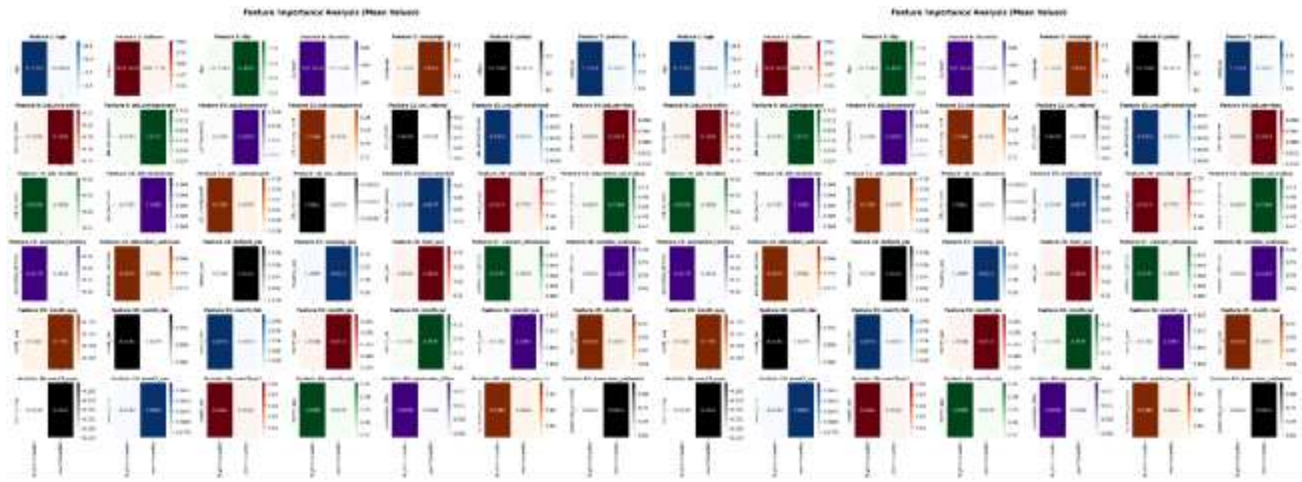
The validity of the clustering model was assessed through sensitivity analysis and precision: 87%, recall: 85%), replicating the results across data changes. The two-cluster solution produced by the X-means clustering method corresponds with clearly delineated behavioral patterns of the population of interest, allowing for a practical basis for strategic decision-making.

SVM result:

Figure 5 presents the mean values of multiple customer attributes corresponding to high and low levels of interaction with Bank Mellat.

Figure 5.

Mean Values of Customer Features Based on Interaction Levels with Bank Mellat



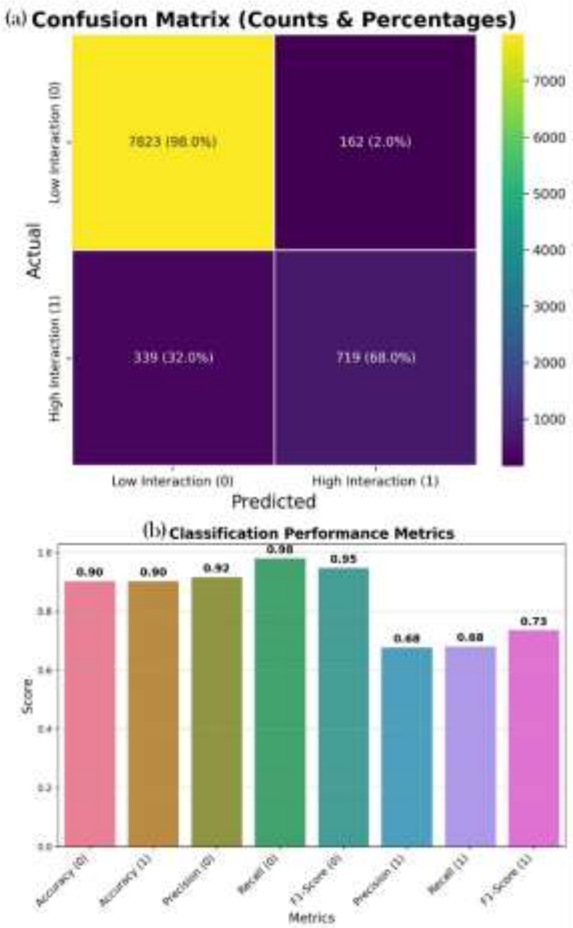
Each subplot features separate customer attributes, including age, balance, occupation type (e.g. blue collar, entrepreneur, management), marital status, educational background, and method of contact. The x-axis charted each subplot included ranges of values or categories for each specific attribute, while the y-axis indicated mean values of those attributes. The variation in color in each subplot highlights the differences between variable distributions. This visualization is beneficial to understand customer attributes that are relevant for predicting high and low interaction levels with the bank.

The figure indicates that duration (537.22s vs. 221.18s), balance (1801.27 vs. 1303.75), and previous interactions (1.17 vs. 1.02) rank amongst the highest differentiating features by high interaction customers and low interaction customers. Similarly, certain job types including entrepreneurship (0.0323 vs. 0.0342) and blue-collar work (0.1339 vs. 0.7260) show meaningful differences as well. Essentially, our findings suggest that financial status, previous interactions, and demographics are among the primary influence on customer interaction, which is valuable information for designing targeted banking models.

Figure 6 summarizes the findings for Bank Mellat customers who demonstrated either a high or low level of interaction based on the features of account balance, call duration, prior interactions, and job type using a Support Vector Machine (SVM) model. The data was split, with 80% used for training and 20% used for testing, with the results reported being generated from the test data.

Figure 6.

Classification Results of Bank Mellat Customers with High and Low Interaction Using SVM



A confusion matrix provides a comprehensive comparison of observed classifications against predicted classifications. The matrix is organized according to the predicted classes and provides important insights regarding correct classifications and misclassifications for each intent subclass. Performance metrics can accompany confusion matrices. Performance measures include accuracy, precision, recall, and F1 score at a 95% confidence level. Each of these measures either communicates performance in general (accuracy) or for a specific subclass of intent (precision, recall, F1 score) while also considering tradeoffs in misclassifications. Continuing with customer intent classification, the confusion matrix in Figure 4 summarizes the classification performance of the AI / ML model. The numbers on the diagonal are correct classifications (correctly classified customers). 98% (7823) of low-intent customers were predicted correctly, and 68% (719) of high intent customers were correctly classified. Misclassifications of high-intent customers included 339 (32%) incorrectly predicted low-intent customers. Misclassifications of low-intent customers included 162 (2%) incorrectly predicted high-intent customers. bar chart in Figure 5 highlights key evaluation metrics. The overall accuracy of the model is 90%. The recall for low-intent customers (0.98) is much higher than high-intent customers (0.68). In this case, the model does well to capture low-intent customers but does moderately well to capture high-intent customers. The precision metric, which reflects true positive cases among positive predictions indicate, for high-intent customers, a precision of 0.68 relative to a low-intent confidence of 0.98. The F1 score, a harmonic mean of precision and recall, indicates a balanced precision and recall measure relative to the high-intent class (F1 = 0.73),suggesting balanced precision and recall for this class. The classification results demonstrate strong performance in identifying low-interaction customers, as evidenced by the high recall (0.98) for this class. On the other hand, the model showed moderate performance in identifying high-interaction customers (recall: 0.68), indicating that feature selection, data balancing techniques, or another model may improve the classification accuracy for this class.

Discussion and Conclusion

The findings of this study validate the capacity of machine learning (ML) methods—specifically k-means clustering and Support Vector Machine (SVM) classification—to enhance B2B customer segmentation in the banking sector. By analyzing a dataset of over 45,000 Bank Mellat clients, the hybrid approach revealed two primary customer clusters: a younger, low-engagement group with lower account balances and fewer interactions, and an older, high-engagement group with greater financial activity, higher balances, and deeper relational ties. The accuracy of the clustering process reached 89%, while the SVM classifier achieved 90% accuracy in predicting customer engagement levels. These results underline the potential of integrating unsupervised and supervised ML techniques for both descriptive and predictive analytics in customer segmentation. More importantly, embedding sustainability considerations within this framework offers a pathway for banks to optimize resources, strengthen loyalty, and design long-term sales strategies.

The behavioral and demographic distinctions identified between the two clusters carry critical implications. The low-engagement group (Cluster 0) was characterized by shorter call durations, infrequent bank interactions, and lower average account balances. This segment appears more transactional in nature and represents customers whose relationships with the bank remain relatively shallow. In contrast, the high-engagement group (Cluster 1) demonstrated longer call durations, more frequent contact with marketing campaigns, and significantly higher account balances. These customers are more relationally oriented, indicating loyalty and sustained use of banking products. The classification performance of the SVM confirms that such behavioral and financial markers can be reliably used to predict future engagement, although the model displayed stronger sensitivity for low-engagement clients. This pattern resonates with prior findings that ML models often capture transactional simplicity with greater precision than relational complexity [22].

The application of k-means clustering to identify homogeneous customer groups aligns with several studies emphasizing the relevance of clustering for customer segmentation. Previous research has demonstrated that clustering enables firms to discover latent patterns in customer data, thereby supporting marketing strategies and increasing customer lifetime value [31, 32]. Our results confirm the ability of clustering to uncover actionable customer profiles and reinforce the observation that clustering methods outperform traditional segmentation approaches such as standard RFM analysis, which tend to overlook multidimensional behavioral features [11]. By adopting clustering, this study extends previous findings by illustrating how segmentation can be paired with sustainability-driven metrics, such as resource optimization and reduced physical interactions, thereby integrating profitability with long-term customer relationship management [3, 5].

The use of SVM classification further expands the methodological contribution of this research. SVM achieved 90% accuracy in classifying customer engagement levels, with particularly strong recall for low-engagement clients. This echoes findings by other scholars who have highlighted the effectiveness of supervised ML algorithms in predicting customer churn and engagement behaviors [17, 26]. However, the moderate recall rate for high-engagement customers underscores the challenge of modeling complex, loyalty-driven interactions. This limitation reflects prior observations that SVM, while robust for binary classification, can struggle when confronted with highly heterogeneous and nonlinear customer behaviors [10]. Such complexity may be better addressed by ensemble methods or deep learning approaches, which have shown promise in managing unstructured or multi-dimensional datasets [7, 21].

In contextualizing these findings, it is important to highlight the significance of sustainability within the segmentation framework. By differentiating customers not only on transactional patterns but also on relational and behavioral engagement, the analysis provides a lens to design sustainable strategies that reduce waste and optimize resource allocation. For example, Cluster 0 customers could be targeted with digital engagement campaigns to minimize physical banking interactions, thereby lowering operational costs and environmental impact. Conversely, Cluster 1 customers justify more resource-intensive loyalty programs due to their demonstrated profitability and sustained relational depth. This interpretation is consistent with the Triple Bottom Line (TBL) perspective, which requires that firms balance financial performance with social and environmental considerations [3]. The results also support prior research highlighting that sustainable segmentation can promote both efficiency and customer loyalty [2, 4].

The integration of ML segmentation with sustainability also contributes to the theoretical discourse on the Resource-Based View (RBV). RBV suggests that firms gain competitive advantage from valuable, rare, inimitable, and organizationally embedded resources [16]. By deploying advanced ML analytics, banks are effectively leveraging unique resources—namely their data and analytical capabilities—to generate strategic insights that are difficult for competitors to replicate. The hybrid k-means and SVM approach thus represents a form of data-driven strategic capability, translating raw data into actionable knowledge that sustains competitive positioning. This contribution resonates with earlier works that conceptualized ML adoption as a strategic resource for improving decision-making and performance [6, 14, 15].

Comparing our results with earlier segmentation frameworks provides further clarity. Traditional RFM and firmographic-based methods, while useful, often fail to capture the dynamic nature of customer interactions [12, 13]. Our approach advances these models by integrating real-time behavioral data such as call duration, campaign interactions, and account balance changes, which provide richer indicators of engagement. In addition, the accuracy achieved by the hybrid model exceeds that reported in studies employing clustering alone [9, 34]. By validating results with sensitivity analysis, the current research also addresses limitations in earlier studies that lacked statistical rigor in cluster determination [20]. This methodological robustness contributes to a more reliable and replicable framework for B2B segmentation.

Our results also align with broader empirical findings in adjacent fields. For instance, Han et al. demonstrated that environmental factors in livestream sales environments, analyzed through ML, significantly impact customer engagement and purchasing decisions [23]. Similarly, Horng and Yenradee highlighted that ML-driven delivery systems improve customer satisfaction and operational efficiency in SMEs [24]. These parallels reinforce the idea that ML segmentation is not confined to banking but represents a cross-sectoral strategic capability that can enhance customer understanding and sustainability outcomes.

Finally, the study provides practical insights for B2B marketing strategies. The low-engagement cluster reflects younger, less financially active customers who may require digital education, simplified financial tools, and personalized digital interactions to increase engagement. The high-engagement cluster, by contrast, represents older, wealthier clients who respond positively to relationship-based strategies such as tailored investment opportunities and loyalty programs. This dichotomy supports findings that effective segmentation requires differentiated engagement strategies, with resource allocation guided by predictive analytics [18, 19, 25]. In this way, our results illustrate the tangible impact of ML segmentation on strategic sales and customer relationship management in banking.

While the study demonstrates the value of combining k-means clustering with SVM classification, several limitations must be acknowledged. First, the reliance on a single dataset from Bank Mellat may limit the generalizability of the results to other banks or industries. Although the dataset was extensive, the behavioral and demographic features used may not capture all relevant dimensions of engagement. Second, the moderate recall for high-engagement customers indicates that the model may underperform when dealing with complex relational dynamics, suggesting the need for more advanced modeling techniques such as ensemble learning or deep neural networks. Third, while the study integrates sustainability considerations conceptually, the operationalization of sustainability metrics was limited to resource efficiency and digital engagement. Broader sustainability indicators, such as environmental footprints of transactions or social impact measures, were not incorporated.

Future research could address these limitations in several ways. Expanding the dataset to include multiple banks or industries would test the external validity of the hybrid approach. Incorporating additional sustainability indicators, such as carbon emissions from physical branch usage or digital inclusivity metrics, could further align segmentation with Triple Bottom Line objectives. Methodologically, future studies might explore hybrid ML frameworks that integrate deep learning, reinforcement learning, or ensemble methods to improve classification performance for complex, high-engagement customers. Additionally, longitudinal studies could assess how customer segments evolve over time, offering insights into the dynamics of loyalty, churn, and sustainable engagement.

From a practical standpoint, banks and B2B organizations should consider implementing hybrid ML models to enhance segmentation accuracy and predictive power. For low-engagement customers, digital-first strategies—including mobile banking tools, automated advisory services, and targeted digital campaigns—can boost engagement while reducing operational costs. For high-engagement customers, banks should prioritize loyalty-building measures such as exclusive services, personalized investment options, and long-term relational programs. Importantly, embedding sustainability into these strategies—by reducing physical resource use, optimizing energy-efficient interactions, and promoting responsible financial products—will ensure that segmentation supports both profitability and broader sustainability goals.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants. Written consent was obtained from all participants in the study.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

[1] M. Wu, P. Andreev, M. Benyoucef, and D. Hood, "Unlocking B2B buyer intentions to purchase: Conceptualizing and validating inside sales purchases," *Decision Support Systems,* 2024, doi: 10.1016/j.dss.2023.114165.

[2] M. Vieth, "Customer segmentation in B2B markets: the relationship between customer segmentation and market orientation," 2018.

[3] J. Elkington, *Cannibals With Forks: The Triple Bottom Line of 21st Century Business*. Oxford: Capstone, 1997.

[4] S. Flambard-Ruaud, "Relationship Marketing: An Innovation in Marketing Theory and Practice," 2015, doi: 10.1007/978-3-319-11845-1_70.

[5] M. Hitka *et al.*, "Sustainability in Marketing through Customer Relationship Management in a Telecommunication Company," *Molecular Microbiology,* 2019, doi: 10.21272/MMI.2019.4-16.

[6] A. Miklosik, M. Kuchta, N. Evans, and S. Zak, "Towards the Adoption of Machine Learning-Based Analytical Tools in Digital Marketing," *IEEE Access,* 2019, doi: 10.1109/ACCESS.2019.2924425.

[7] B. H. Chen *et al.*, "Uncloaking hidden repeating fast radio bursts with unsupervised machine learning," *Monthly Notices of the Royal Astronomical Society,* vol. 509, no. 1, pp. 1227-1236, 2022, doi: 10.1093/mnras/stab2994.

[8] S. N. Lathifah and Z. F. Azzahra, "AI-Driven Customers Segmentation Using K-Means Clustering," *G-Tech: Jurnal Teknologi Terapan,* vol. 9, no. 1, pp. 320-329, 2025, doi: 10.70609/gtech.v9i1.6202.

[9] A. Z. P. Aufa Zahrani Putri, "Penerapan Algoritma Fuzzy C-Means Pada Segmentasi Pelanggan B2B dengan Model LRFM," *Jurnal Media Informatika Budidarma,* vol. 7, no. 3, 2023, doi: 10.30865/mib.v7i3.6150.

[10] S. Mishra, P. Nayak, R. K. Mallick, D. A. Gadanayak, and G. Panda, "PQ event identification in PV-wind based distribution network with variational mode decomposition and novel feature enabled random forest classifier," *International Journal of Emerging Electric Power Systems,* vol. 25, no. 3, pp. 393-404, 2024, doi: 10.1515/ijeeps-2023-0123.

[11] M. Chattopadhyay and S. K. Mitra, "Elucidating strategic patterns from target customers using multi-stage RFM analysis," *Journal of Global Scholars,* 2023, doi: 10.1080/21639159.2022.2080094.

[12] A. B. Hadid, S. Bouguelia, and H. Kheddouci, "A New Method of B2B Customer Segmentation Based on Firmographic Data, and RFM and Graph Models," in *2024 IEEE International Conference on e-Business Engineering*, 2024, pp. 81-86, doi: 10.1109/ICEBE62490.2024.00021.

[13] P. Li, C. Wang, J. Wu, and R. Madleňák, "An E-commerce customer segmentation method based on RFM weighted K-means," in *2022 International Conference on Management Engineering, Software Engineering and Service Sciences*, 2022, pp. 61-68, doi: 10.1109/ICMSS55574.2022.00017.

[14] A. S. A. Alwabel and X. J. Zeng, "Data-driven modeling of technology acceptance: A machine learning perspective," *Expert Systems With Applications,* 2021, doi: 10.1016/j.eswa.2021.115584.

[15] T. Dahlberg and T. Nokkala, "A framework for the corporate governance of data - theoretical background and empirical evidence," 2015, doi: 10.3846/bme.2015.254.

[16] J. B. Barney, "Firm resources and sustained competitive advantage," *Journal of Management,* vol. 17, no. 1, pp. 99-120, 1991, doi: 10.1177/014920639101700108.

[17] O. Wisesa, A. Andriansyah, and O. I. Khalaf, "Prediction Analysis for Business To Business (B2B) Sales of Telecommunication Services using Machine Learning Techniques," *Majlesi Journal of Electrical Engineering,* 2020, doi: 10.29252/MJEE.14.4.145.

[18] M. R. A. Purnomo, A. Azzam, and A. U. Khasanah, "Effective Marketing Strategy Determination Based on Customers Clustering Using Machine Learning Technique," 2020, doi: 10.1088/1742-6596/1471/1/012023.

[19] L. Abidar, D. Zaidouni, and A. En-Nouaary, "Customer Segmentation With Machine Learning: New Strategy For Targeted Actions," in *International Conference on Intelligent Systems: Theories and Applications*, 2020, doi: 10.1145/3419604.3419794.

[20] N. Sakina, A. P. Arun, and P. K. Gupta, "Optimizing Customer Segmentation: A Comparative Analysis of Clustering Algorithms Using Evaluation Metrics," in *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions*, 2024, pp. 1-6, doi: 10.1109/CSITSS64042.2024.10816952.

[21] K. K. Agrawal and G. Agarwal, "A Comparative Study of Deep Learning vs. Machine Learning Algorithms for Brain Tumor Detection," in *2024 1st International Conference on Advances in Computing, Communication and Networking*, 2024, pp. 1001-1005, doi: 10.1109/ICAC2N63387.2024.10894885.

[22] A. Manzoor, M. A. Qureshi, E. Kidney, and L. Longo, "A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners," *IEEE Access,* 2024, doi: 10.1109/ACCESS.2024.3402092.

[23] L. Han, J. Fang, Q. Zheng, B. T. George, and M. Liao, "Unveiling the effects of livestream studio environment design on sales performance: A machine learning exploration," *Industrial Marketing,* 2024, doi: 10.1016/j.indmarman.2023.12.021.

[24] S. Horng and P. Yenradee, "Delivery Service Management System Using Google Maps for SMEs in Emerging Countries," *Computers, Materials & Continue,* 2023, doi: 10.32604/cmc.2023.038764.

[25] M. O'Brien, Y. Liu, H. Chen, and R. F. Lusch, "Gaining insight to B2B relationships through new segmentation approaches: Not all relationships are equal," *Expert Systems With Applications,* 2020, doi: 10.1016/j.eswa.2020.113767.

[26] A. Sheikh, T. Ghanbarpour, and D. Gholamiangonabadi, "A preliminary study of fintech industry: a two-stage clustering analysis for customer segmentation in the B2B setting," 2019, vol. 26, 2 ed., pp. 197-207, doi: 10.1080/1051712X.2019.1603420.

[27] S. GhGolamveisy *et al.*, "Application of data mining technique for customer purchase behavior via Extended RFM model with focus on BCG matrix from a data set of online retailing," *Journal of Infrastructure, Policy and Development,* vol. 8, no. 7, p. 4426, 2024, doi: 10.24294/jipd.v8i7.4426.

[28] S. Sancar and M. Uzun-Per, "Feature Selection in Customer Churn Analysis: Case Study in B2B Business," in *IEEE International Conference on E-Business Engineering*, 2022, doi: 10.1109/ICEBE55470.2022.00053.

[29] A. B. Madeira, J. A. G. d. Silveira, and L. A. Toledo, "Marketing Segmentation: Your Role For Diversity in Dynamical Systems," *GESTÃO.Org: Revista Eletrônica de Gestão Organizacional,* 2015.

[30] S. Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods," in *International Conference on Artificial Intelligence*, 2018, doi: 10.1109/IDAP.2018.8620892.

[31] I. Fuentes, I. Fuentes, G. Nápoles, L. Arco, L. Arco, and K. Vanhoof, *Customer Segmentation Using Multiple Instance Clustering and Purchasing Behaviors*. 2018.

[32] D. A. Kandeil, A. A. Saad, and S. M. Youssef, "A two-phase clustering analysis for B2B customer segmentation," in *2014 International Conference on Intelligent Networking and Collaborative Systems*, 2014, pp. 221-228, doi: 10.1109/INCoS.2014.49.

[33] W. Zhu, N. Zeng, and N. Wang, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations," *NESUG Proceedings: Health Care and Life Sciences,* p. 67, 2010.

[34] N. R. Maulina, I. Surjandari, and A. M. M. Rus, "Data Mining Approach for Customer Segmentation in B2B Settings using Centroid-Based Clustering," in *International Conference on Service Systems and Service Management*, 2019, doi: 10.1109/ICSSSM.2019.8887739.