Management Strategies and Engineering Sciences 2026; 8(6):1-11

Development of Machine Learning-Based Customer Credit Scoring Models: Analysis of Bank Melli Iran Data for Credit Risk Prediction

Behzad Abdi[1], Afshin Baghfalaki[2*] Alireza Moradi[2] Adel Fatemi[3]

[1] Department of Financial Engineering, Ker.C., Islamic Azad University, Kermanshah, Iran
[2] Department of Economics, Ker.C., Islamic Azad University, Kermanshah, Iran
[3] Department of Statistics, Sa.C., Islamic Azad University, Sanandaj, Iran

*** Corresponding author email address**: Afshin.Baghfalaki@iau.ac.ir

Received: 2026-01-01 **Revised:** 2026-04-27 **Accepted:** 2026-05-02 **Initial Publish:** 2026-05-24 **Final Publish:** 2026-11-01

Abstract

The aim of this study was to develop and evaluate customer credit risk prediction models using machine learning algorithms based on real-world data from Bank Melli Iran. In this research, a dataset consisting of 24,860 individual customers with 42 financial, demographic, behavioral, and credit-related variables was utilized. The default rate of the dataset was 14.8%, providing an appropriate basis for evaluating classification models. After conducting preprocessing procedures, including outlier removal, missing value imputation, and data normalization, five models—Logistic Regression, Decision Tree, Random Forest, XGBoost, and Multilayer Perceptron Neural Network—were trained. Model performance was evaluated using indicators such as Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC). The results demonstrated that the XGBoost algorithm achieved the best performance in predicting customer default probability, with an accuracy of 86.1% and an AUC value of 0.912. Variable importance analysis also revealed that the debt-to-income ratio, number of overdue installments, history of payment delay, and average account balance were among the most influential factors in determining credit risk. The findings indicate that machine learning models, particularly gradient boosting-based methods, can significantly enhance the accuracy of banking credit scoring systems.

Keywords: *Customer Credit Scoring, Credit Risk, Machine Learning, Default Prediction, XGBoost Algorithm*

How to cite this article:

Abdi, B., Baghfalaki, A., Moradi, A., & Fatemi, A. (2026). Development of Machine Learning-Based Customer Credit Scoring Models: Analysis of Bank Melli Iran Data for Credit Risk Prediction. Management Strategies and Engineering Sciences, 8(6), 1-11.

1. Introduction

The banking industry has undergone substantial transformation in recent decades due to rapid technological advancement, digitalization of financial services, increasing competition, and the growing complexity of financial markets. In this environment, effective credit risk management has become one of the most critical determinants of banking stability, profitability, and sustainability. Credit risk refers to the probability that borrowers may fail to fulfill their financial obligations according to agreed contractual terms, thereby exposing financial institutions to potential losses. Since lending activities constitute one of the primary sources of revenue generation for commercial banks, accurate assessment of customer creditworthiness has become an essential strategic requirement for banking systems worldwide. The increasing complexity of customer financial behavior and the expansion of banking transactions have intensified the need for advanced analytical tools capable of predicting default risk with higher levels of accuracy and reliability [1, 2].

Traditional banking systems have historically relied on conventional statistical techniques and expert-based judgment for evaluating customer creditworthiness. Although these methods have played an important role in banking decision-making for many years, their predictive capability is often limited when dealing with large-scale, multidimensional, and nonlinear financial data. Conventional credit scoring systems usually depend on a restricted number of financial indicators and are often unable to capture hidden behavioral patterns embedded within customer transaction histories. As banking environments become increasingly data-intensive, traditional approaches face substantial challenges in maintaining predictive accuracy and adaptability under dynamic economic conditions [3, 4]. Consequently, banking institutions are increasingly moving toward intelligent data-driven systems that can process massive datasets and identify complex relationships among financial variables.

Machine learning has emerged as one of the most influential technological innovations in modern financial analytics and banking management. Machine learning algorithms possess the ability to automatically learn patterns from historical data, improve predictive performance over time, and identify nonlinear relationships that may remain undetected through classical statistical approaches. The application of machine learning techniques in banking has expanded rapidly across multiple domains, including fraud detection, customer segmentation, portfolio optimization, marketing analysis, operational risk management, and especially credit risk prediction. Advanced algorithms such as Random Forest, XGBoost, Support Vector Machines, and Neural Networks have demonstrated superior predictive performance compared with conventional regression-based approaches in many financial studies [5, 6].

One of the major advantages of machine learning models in credit scoring systems is their capability to integrate diverse categories of information simultaneously. Modern banking databases contain extensive information regarding customer demographics, account transactions, repayment histories, borrowing patterns, spending behavior, and financial relationships. Machine learning algorithms can analyze these multidimensional data structures and extract meaningful predictive patterns that significantly enhance the precision of default prediction models. Furthermore, ensemble learning approaches and gradient boosting techniques are particularly effective in handling imbalanced datasets and nonlinear interactions among variables, which are common characteristics of credit risk datasets [5, 7].

The importance of effective credit risk assessment extends beyond individual banking institutions and directly influences financial system stability at national and international levels. Weak credit risk management practices may lead to excessive non-performing loans, liquidity shortages, reduced banking profitability, and ultimately financial crises. Recent international banking studies have emphasized that accurate credit risk evaluation contributes significantly to profitability improvement, portfolio efficiency, and institutional resilience in volatile economic conditions [1]. In addition, emerging global challenges such as climate-related financial risks and sustainable investment transitions have further increased the complexity of banking risk management systems. Contemporary research has demonstrated that environmental and macroeconomic variables may indirectly affect banking credit portfolios and borrower repayment behavior [2, 8].

In the context of the Iranian banking system, the issue of customer credit scoring has become increasingly important due to economic fluctuations, inflationary pressures, regulatory changes, and the expansion of retail banking services. Iranian banks face growing pressure to reduce non-performing loans and improve the efficiency of lending decisions. Traditional evaluation methods, which are often heavily dependent on collateral and manual assessments, may no longer provide sufficient predictive accuracy under contemporary financial conditions. Several Iranian studies have therefore attempted to develop localized credit scoring models using statistical and computational approaches. Research conducted on customers of Bank Sepah demonstrated that metaheuristic algorithms can improve prediction accuracy compared with conventional methods [5]. Similarly, studies on Bank Mellat customers indicated that hybrid machine learning models can significantly enhance customer credit scoring systems by integrating multiple analytical approaches [6].

Credit scoring systems also play a strategic role in improving customer relationship management and banking service quality. Accurate identification of customer risk profiles enables banks to allocate financial resources more efficiently, personalize service offerings, and optimize lending strategies according to customer characteristics. Modern customer-oriented banking systems increasingly rely on intelligent analytics to support strategic alliances, customer segmentation, and relationship management processes [9, 10]. In this regard, the integration of machine learning into banking decision-making frameworks can enhance both operational performance and customer satisfaction simultaneously.

Another important aspect of modern banking management is the relationship between service quality perception and customer financial behavior. Studies have shown that customer trust, perceived service quality, and banking experience may indirectly influence repayment behavior and long-term customer loyalty. Improvements in physical banking environments and service quality dimensions contribute to stronger customer relationships and may reduce behavioral risks associated with financial obligations [11]. Moreover, emotional branding and customer experience management have become increasingly important competitive tools in the banking sector, influencing customer retention and financial engagement [12]. Consequently, comprehensive credit scoring systems should not solely focus on traditional financial variables but should also incorporate behavioral and relational dimensions associated with customer banking interactions.

Recent advances in artificial intelligence and financial technology have enabled banks to implement highly sophisticated predictive systems capable of real-time decision-making. Algorithms such as XGBoost have gained considerable attention because of their ability to optimize predictive performance while controlling overfitting through regularization mechanisms. These algorithms can process high-dimensional datasets efficiently and generate feature importance rankings that improve the interpretability of credit scoring systems. Feature importance analysis is particularly valuable in banking applications because it allows financial managers to identify the most influential determinants of customer default behavior and improve risk mitigation policies accordingly [6, 7].

Despite the growing application of machine learning techniques in global banking systems, empirical evidence regarding their implementation within Iranian banking institutions remains relatively limited. Many existing studies in Iran have primarily relied on smaller datasets, econometric models, or limited variable structures. Furthermore, there is insufficient evidence comparing the performance of multiple machine learning algorithms using large-scale real banking data from major Iranian financial institutions. The need for comprehensive comparative analysis is therefore highly significant, particularly for identifying the most effective predictive approaches under local banking conditions [4, 13].

The significance of ethical considerations in customer credit evaluation has also become increasingly prominent in recent banking research. Ethical validation in credit assessment emphasizes fairness, transparency, accountability, and the avoidance of discriminatory lending decisions. Advanced machine learning systems must therefore balance predictive accuracy with ethical risk management principles to ensure sustainable banking operations and regulatory compliance [14]. The integration of explainable machine learning approaches within banking systems may contribute to improved transparency and managerial trust in automated credit decision-making processes.

Moreover, collateral and guarantee management continue to represent fundamental dimensions of banking credit risk assessment. In many banking systems, collateral quality and guarantee structures play an important role in reducing potential financial losses associated with customer default. However, recent studies suggest that relying solely on collateral-based evaluations may be insufficient in highly dynamic financial environments. Integrated risk management frameworks combining behavioral, financial, and guarantee-related indicators are therefore increasingly recommended for improving the effectiveness of banking credit systems [15]. This multidimensional perspective aligns closely with machine learning methodologies capable of integrating heterogeneous information sources into unified predictive frameworks.

Customer segmentation has also emerged as a critical component of intelligent banking systems. Advanced clustering and segmentation algorithms enable banks to classify customers according to behavioral patterns, financial characteristics, and risk profiles. Such segmentation approaches improve the targeting accuracy of banking products and support more efficient allocation of financial resources. Recent studies utilizing improved clustering algorithms have demonstrated substantial improvements in banking marketing strategies and customer management processes [10]. Integrating segmentation insights into credit scoring models may further enhance the precision of risk prediction and portfolio management strategies.

The development of advanced credit risk prediction systems is therefore not merely a technological innovation but also a strategic managerial necessity for modern banking institutions. Accurate prediction of customer default behavior contributes to improved portfolio quality, reduced financial losses, optimized lending decisions, enhanced customer management, and greater institutional resilience. Given the increasing availability of large-scale banking datasets and advancements in artificial intelligence technologies, the implementation of machine learning-based credit scoring systems has become increasingly feasible and economically valuable for financial institutions worldwide [1, 8].

Accordingly, the aim of this study is to develop and evaluate machine learning-based customer credit scoring models using real-world data from Bank Melli Iran in order to identify the most accurate predictive model for customer credit risk assessment and default prediction.

2. Methodology

This study was conducted using a quantitative, applied, and data-driven research design aimed at developing and evaluating machine learning models for customer credit risk prediction. The methodological framework was established based on standard data mining and predictive analytics procedures, including data acquisition, preprocessing, feature categorization, model development, training, validation, and performance evaluation. The research focused on identifying the most accurate predictive model for estimating customer default probability using real-world banking data obtained from Bank Melli Iran.

The statistical population of the study consisted of individual customers who had received financial facilities and loan services from the bank. The dataset included information related to 24,860 customers collected over a three-year period. For each customer, 42 variables were extracted from the banking information system. These variables represented different dimensions of customer characteristics, including demographic, financial, credit-related, and behavioral attributes. The dependent variable of the study was customer default status, which was defined as a binary classification variable. A value of $Y = 1$represented customers who had defaulted on their loan obligations, whereas $Y = 0$represented customers who had repaid their financial obligations regularly and without significant delinquency.

Initial examination of the dataset revealed that approximately 14.8% of customers were classified as default cases, indicating the existence of class imbalance in the data. Although the imbalance level was not considered extreme, it represented a realistic distribution commonly observed in banking credit datasets and therefore provided an appropriate context for evaluating classification algorithms under practical financial conditions. The relatively large sample size and diversity of variables enhanced the robustness and generalizability of the predictive models developed in this study.

The study design emphasized predictive accuracy and comparative evaluation across different machine learning approaches. To ensure methodological consistency, the dataset was divided into separate training and testing subsets. Approximately 70% of the observations, corresponding to 17,402 customer records, were allocated to the training dataset, while the remaining 30%, equivalent to 7,458 records, were reserved for out-of-sample testing and model validation. This partitioning strategy enabled unbiased evaluation of model performance and reduced the risk of overfitting during the training phase.

The data used in this research were extracted from the internal banking databases of Bank Melli Iran and included a comprehensive set of customer-level financial and behavioral indicators. The variables were categorized into four major groups consisting of demographic variables, financial variables, credit variables, and behavioral variables. This multidimensional structure enabled the predictive models to capture various aspects of customer financial behavior and repayment capability.

Demographic variables included characteristics such as customer age, gender, marital status, and educational level. These variables were incorporated because demographic characteristics are frequently associated with variations in financial responsibility, consumption behavior, and repayment stability. Financial variables included monthly income, average account balance, debt-to-income ratio, and other indicators associated with customer liquidity and financial strength. Among these variables, the debt-to-income ratio represented one of the most critical predictors of default risk because it reflected the extent of customer leverage relative to repayment capacity.

Credit-related variables included the number of previous loans, repayment history, delayed payment records, and the number of overdue installments. These variables provided direct evidence regarding the historical creditworthiness of customers and their previous interactions with the banking system. Behavioral variables included transaction frequency, average account turnover, cash withdrawal patterns, and account usage behavior. Such variables enabled the identification of latent behavioral patterns that may not be directly observable through traditional financial indicators alone.

Before model implementation, the raw data underwent several preprocessing procedures to improve data quality and ensure compatibility with machine learning algorithms. In the first stage, records containing more than 30% missing values were removed from the dataset to prevent distortion in model training. For the remaining missing values, statistical imputation techniques were applied. Numerical variables were replaced using mean imputation, whereas categorical variables were completed using mode substitution. This procedure minimized information loss while maintaining statistical consistency within the dataset.

Following data cleaning, numerical variables were standardized using the Z-score normalization method in order to eliminate scale-related distortions among variables. The standardization formula was defined as:

$$Z = \frac{X - \mu}{\sigma}$$

where:

X = Observed value of the variable
μ = Mean of the variable
σ = Standard deviation of the variable

Standardization ensured that variables with larger numerical scales did not dominate the learning process of the algorithms, particularly in models sensitive to feature magnitude such as Logistic Regression and Neural Networks. In addition, outlier observations were identified and removed using statistical threshold techniques to reduce noise and improve predictive stability.

The data analysis process involved the implementation and comparison of five predictive models, including Logistic Regression, Decision Tree, Random Forest, XGBoost, and Multilayer Perceptron Neural Network models. These algorithms were selected due to their widespread application in financial risk modeling and their ability to capture both linear and nonlinear relationships among variables.

Logistic Regression was employed as the baseline statistical classification model because of its interpretability and extensive use in credit scoring systems. In this model, the probability of customer default was estimated using the sigmoid function:

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}}$$

where:

$P(Y = 1 \mid X)$ = Probability of customer default
β_0 = Intercept term
β_i = Regression coefficients
X_i = Independent variables

Decision Tree analysis was utilized to model nonlinear decision boundaries through hierarchical data partitioning. The model recursively divided observations into homogeneous subsets in order to maximize class separation between defaulting and non-defaulting customers. The splitting criterion was based on the Gini impurity index defined as:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2$$

where:

p_i = Probability of belonging to class i
C = Number of classes

The Random Forest model was implemented as an ensemble learning approach combining multiple decision trees generated from random subsets of the data. The final prediction was obtained through aggregation across all trees according to:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b(X)$$

where:

B = Number of trees
$T_b(X)$ = Prediction generated by tree b

The XGBoost algorithm was applied as an advanced gradient boosting technique designed to optimize predictive accuracy while controlling model complexity. The objective function minimized by the algorithm was expressed as:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where the regularization component was defined as:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2$$

In this formulation:

T = Number of leaves in the decision tree
w_j = Weight assigned to each leaf
γ and λ
= Regularization parameters controlling model complexity

The Multilayer Perceptron Neural Network was employed to identify highly complex and nonlinear patterns in customer credit behavior. The neural network architecture consisted of interconnected hidden layers and nonlinear activation functions enabling deep feature extraction and adaptive learning.

Model performance was evaluated using multiple classification metrics, including Accuracy, Precision, Recall, F1-Score, and Area Under the Receiver Operating

Characteristic Curve (AUC). These evaluation indicators provided a comprehensive assessment of classification quality and discriminatory power. The evaluation metrics were calculated using the confusion matrix consisting of True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN).

Recall was calculated as:

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

Precision was calculated as:

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

Accuracy was calculated as:

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

The F1-Score represented the harmonic mean of Precision and Recall, while the AUC metric evaluated the discriminatory capability of each model in distinguishing default and non-default classes across different threshold levels. Collectively, this analytical framework enabled a comprehensive comparison of machine learning techniques for customer credit risk prediction and facilitated the identification of the most efficient predictive model for banking credit scoring applications.

3. Findings and Results

In this section, the results obtained from implementing different machine learning models for customer credit risk prediction are presented. First, the performance of the models is compared based on standard evaluation metrics. Subsequently, the confusion matrix and discriminatory power of the models are examined. Finally, the most influential variables affecting customer credit risk are analyzed.

After training the models using the training dataset, their performance was evaluated on the testing dataset. The evaluation metrics included Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC). The results indicated that advanced machine learning models outperformed the classical statistical model represented by Logistic Regression.

Table 1. Comparison of Credit Risk Prediction Model Performance

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.782	0.701	0.653	0.676	0.801
Decision Tree	0.801	0.732	0.681	0.705	0.824
Random Forest	0.842	0.781	0.734	0.757	0.889
XGBoost	0.861	0.804	0.759	0.781	0.912
Neural Network (MLP)	0.847	0.789	0.741	0.764	0.896

The results presented in Table 1 demonstrate that the XGBoost model achieved the best overall performance among the examined models, with an accuracy of 86.1% and an AUC value of 0.912. Following XGBoost, the Random Forest and Multilayer Perceptron Neural Network models ranked second and third, respectively. These findings are consistent with international studies indicating that gradient boosting algorithms provide highly effective performance in credit scoring and financial risk prediction problems.

To evaluate the predictive capability of the best-performing model more precisely, the confusion matrix of the XGBoost model was calculated. The confusion matrix illustrates the extent to which the model successfully identified defaulting and non-defaulting customers.

Table 2. Confusion Matrix of the XGBoost Model

	Actual Default	Actual Non-Default
Predicted Default	1,692	412
Predicted Non-Default	356	4,998

Based on Table 2, the performance indicators of the model were calculated as follows:

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

$$Recall = \frac{1692}{1692 + 356} = 0.826$$

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

$$Precision = \frac{1692}{1692 + 412} = 0.804$$

These results indicate that the model possesses a strong capability for identifying high-risk customers and can successfully predict a substantial proportion of potential default cases. The relatively high Recall value demonstrates the effectiveness of the model in minimizing false negative classifications, which is critically important in banking risk management systems.

To evaluate the discriminatory power of the models, the Receiver Operating Characteristic (ROC) curve was analyzed. The Area Under the Curve (AUC) represents the ability of the model to distinguish between default and non-default classes. The AUC value ranges between 0 and 1, and values closer to 1 indicate superior classification performance.

Table 3. Comparison of AUC Values Across Models

Model	AUC Value
Logistic Regression	0.801
Decision Tree	0.824
Random Forest	0.889
XGBoost	0.912
Neural Network	0.896

The results presented in Table 3 reveal that all machine learning models outperformed the Logistic Regression model in terms of discriminatory capability. Among the examined models, XGBoost achieved the highest AUC value, indicating the strongest ability to differentiate between defaulting and non-defaulting customers. The Random Forest and Neural Network models also demonstrated strong discriminatory performance, further confirming the effectiveness of ensemble and deep learning approaches in credit risk prediction tasks.

One of the primary advantages of tree-based models is the ability to extract feature importance scores. In this study, feature importance was calculated based on the XGBoost model in order to identify the variables with the greatest influence on customer credit risk prediction.

Table 4. Most Important Variables Affecting Credit Risk

Rank	Variable	Importance Score
1	Debt-to-Income Ratio	0.184
2	Number of Overdue Installments	0.167
3	History of Payment Delay	0.151
4	Average Account Balance	0.128
5	Number of Previous Loans	0.102
6	Average Account Turnover	0.094
7	Customer Age	0.082
8	Monthly Income	0.071

The findings shown in Table 4 indicate that the debt-to-income ratio and the number of overdue installments were the most influential determinants of customer credit risk. This finding is consistent with banking risk management theories, as high debt levels generally reflect greater financial pressure on customers and consequently increase the probability of loan default. Furthermore, variables associated with repayment behavior and account management patterns demonstrated substantial predictive importance, emphasizing the significance of behavioral financial indicators in modern credit scoring systems.

To summarize the findings, the models were compared based on predictive accuracy, discriminatory power, and capability to identify high-risk customers.

Table 5. Final Ranking of the Models

Rank	Model	Result
1	XGBoost	Best Performance
2	Random Forest	Very Strong Performance
3	Neural Network	Appropriate Performance
4	Decision Tree	Moderate Performance
5	Logistic Regression	Baseline Model

The overall findings demonstrate that the use of advanced machine learning algorithms can significantly improve the accuracy of credit risk prediction systems. Among the examined models, the XGBoost algorithm provided the best overall performance and may therefore serve as an effective tool for banking credit scoring and financial risk management systems. The superior performance of ensemble learning and gradient boosting techniques further highlights the importance of adopting intelligent analytical systems within modern banking infrastructures.

4. Discussion and Conclusion

The findings of this study demonstrated that machine learning algorithms significantly improved the prediction accuracy of customer credit risk compared with traditional statistical methods. Among the examined models, the XGBoost algorithm achieved the highest predictive performance, with an accuracy of 86.1% and an AUC value of 0.912, outperforming Logistic Regression, Decision Tree, Random Forest, and Multilayer Perceptron Neural Network models. These results indicate that advanced ensemble learning techniques possess superior capability in identifying complex and nonlinear relationships among customer financial, behavioral, and demographic variables. The findings are consistent with previous studies emphasizing the effectiveness of machine learning algorithms in banking credit scoring systems [5, 6]. The superior performance of XGBoost can be attributed to its gradient boosting structure, regularization mechanisms, and iterative optimization process, which collectively improve model generalization and reduce overfitting problems commonly observed in financial datasets.

The findings also revealed that traditional Logistic Regression demonstrated the weakest performance among the examined models. Although Logistic Regression has long been regarded as a foundational method in banking credit analysis because of its interpretability and simplicity, the lower predictive accuracy observed in this study highlights its limitations when dealing with multidimensional and nonlinear banking data structures. Modern banking environments generate highly complex transactional and behavioral information that often cannot be adequately modeled through linear statistical assumptions alone. Similar conclusions were reported in studies conducted on Iranian banking customers, which indicated that conventional econometric and regression-based approaches may provide lower predictive efficiency compared with intelligent computational techniques [3, 7]. Therefore, the current findings further reinforce the argument that banking institutions should gradually transition from purely traditional scoring systems toward hybrid and machine learning-driven analytical infrastructures.

Another important finding of the study was the strong discriminatory power of the XGBoost model, reflected in the high AUC value and confusion matrix indicators. The Recall value of 0.826 demonstrated that the model successfully identified a substantial proportion of customers who were likely to default. This issue is particularly important in banking systems because false negative classifications may lead to significant financial losses through the approval of high-risk customers. The Precision value of 0.804 also indicated that the model maintained acceptable accuracy in predicting actual default cases, thereby reducing the likelihood of incorrectly classifying low-risk customers as high-risk borrowers. These results suggest that advanced machine learning models can provide an effective balance between risk sensitivity and predictive precision. Similar observations have been reported in studies examining hybrid and metaheuristic approaches for customer credit assessment in Iranian banking systems [5, 13].

The findings additionally demonstrated that Random Forest and Multilayer Perceptron Neural Network models achieved strong predictive performance and ranked immediately after XGBoost. This result confirms the broader effectiveness of ensemble learning and deep learning methods in banking risk prediction tasks. Random Forest benefited from aggregation across multiple decision trees, which improved predictive stability and reduced variance. Likewise, the neural network model demonstrated considerable capability in identifying hidden nonlinear patterns among variables. However, despite their strong performance, both models remained slightly weaker than XGBoost in terms of overall accuracy and discriminatory power. This finding aligns with previous studies suggesting that gradient boosting techniques often outperform other machine learning approaches in structured tabular financial datasets because of their sequential learning architecture and optimization mechanisms [6, 11].

One of the most significant contributions of this study lies in the analysis of feature importance within the XGBoost framework. The results indicated that the debt-to-income ratio, number of overdue installments, payment delay history, and average account balance were among the most influential variables affecting customer credit risk. The dominant role of the debt-to-income ratio is theoretically consistent with banking risk management principles because customers with excessive debt burdens are naturally more vulnerable to repayment difficulties and financial distress. This finding is highly consistent with prior studies emphasizing the critical role of financial leverage indicators in determining borrower creditworthiness [14, 15]. Customers with high debt obligations relative to their income capacity are more likely to encounter liquidity shortages, which substantially increases the probability of default.

The importance of overdue installments and payment delay history also highlights the strong predictive value of historical repayment behavior in banking systems. Behavioral finance theories suggest that past financial behavior often serves as one of the most reliable indicators of future financial discipline and repayment reliability. The current findings therefore support the argument that customer behavioral variables should be integrated more comprehensively into modern credit scoring systems. Similar results were observed in studies conducted on Iranian bank customers, where repayment history and delinquency indicators significantly influenced customer ranking and credit evaluation outcomes [4, 13]. The inclusion of behavioral indicators within machine learning models enhances the ability of banks to identify latent patterns of financial irresponsibility that may not be immediately observable through traditional financial statements alone.

The findings further demonstrated the importance of average account balance and account turnover variables in predicting customer default risk. Customers maintaining stable account balances and regular transaction flows are generally more financially stable and less likely to experience repayment problems. These variables may also indirectly reflect customer engagement with the banking system and overall financial discipline. Recent banking management studies have similarly emphasized the growing importance of customer behavioral analytics and transactional patterns in modern financial decision-making processes [10, 12]. The integration of transaction-based behavioral data into predictive models therefore represents an important advancement in intelligent banking systems.

Another notable implication of the findings concerns the strategic role of machine learning in improving banking profitability and operational efficiency. Effective credit risk prediction contributes directly to reducing non-performing loans, improving portfolio quality, and optimizing resource allocation. International evidence has demonstrated that stronger credit risk management practices positively influence banking profitability and institutional stability [1]. By reducing lending errors and improving customer classification accuracy, machine learning systems can help banks minimize financial losses associated with risky lending activities while simultaneously increasing operational productivity.

The results of this study also have important implications for customer relationship management and service personalization within banking institutions. Intelligent credit scoring systems enable banks to classify customers more accurately according to financial behavior, risk levels, and borrowing characteristics. Such segmentation supports the development of personalized financial services and targeted marketing strategies. Previous studies have shown that customer-oriented banking systems increasingly rely on analytical technologies to improve customer satisfaction, service quality, and strategic decision-making [9, 10]. Therefore, the integration of machine learning into banking operations may generate both financial and relational benefits for banking institutions.

Furthermore, the findings support the growing argument that modern banking systems should adopt multidimensional and integrated approaches to risk management. Contemporary banking risk is no longer solely determined by traditional financial variables but is also influenced by behavioral, environmental, and macroeconomic conditions. Recent international research has highlighted the relationship between broader economic dynamics, environmental transitions, and banking credit risk exposure [2, 8]. Consequently, intelligent machine learning frameworks capable of integrating heterogeneous data sources may become increasingly important for sustainable banking management in future financial systems.

The ethical dimension of credit scoring systems should also be considered when interpreting the findings of this study. Although machine learning algorithms improve predictive performance substantially, banking institutions must ensure that automated decision-making systems operate transparently and fairly. Ethical validation of customer assessment systems has become a major concern in contemporary banking research because biased algorithms may unintentionally produce discriminatory lending outcomes [14]. Therefore, while implementing machine learning systems, banks should simultaneously develop transparent governance frameworks and explainable artificial intelligence mechanisms to maintain customer trust and regulatory compliance.

The findings additionally confirm the growing importance of technological transformation in the Iranian banking sector. As banking competition intensifies and customer expectations continue to evolve, financial institutions increasingly require intelligent analytical infrastructures capable of processing large-scale financial data efficiently. The successful performance of advanced machine learning models in this study demonstrates that Iranian banking institutions possess substantial opportunities to modernize their credit evaluation systems through artificial intelligence technologies. Similar trends have been emphasized in previous Iranian banking studies focusing on customer scoring and risk management modernization [7, 16]. The implementation of such systems may improve both institutional competitiveness and financial stability within the banking sector.

One limitation of this study is that the dataset was restricted to customers of a single banking institution, which may reduce the generalizability of the findings to other banks or financial environments. In addition, the study focused primarily on structured banking data and did not incorporate alternative data sources such as social media activity, mobile banking behavior, or macroeconomic indicators. Another limitation is that the models were evaluated using historical data collected over a fixed time period, whereas customer financial behavior may evolve dynamically under changing economic conditions.

Future research may extend the current framework by incorporating larger and more diverse datasets obtained from multiple banking institutions and financial sectors. Researchers may also examine the integration of unstructured data sources, including text mining, mobile transaction data, and customer digital behavior, to improve predictive performance further. Comparative analysis of explainable artificial intelligence methods and fairness-aware machine learning algorithms may additionally contribute to the development of more transparent and ethically responsible credit scoring systems.

From a practical perspective, banking institutions should gradually integrate advanced machine learning algorithms into their credit risk management infrastructures to improve lending accuracy and reduce non-performing loans. Banks should also invest in data governance systems, employee training, and artificial intelligence capabilities to support the effective implementation of intelligent analytical systems. Furthermore, policymakers and banking regulators should establish supervisory frameworks that ensure transparency, fairness, and accountability in automated credit scoring systems while encouraging technological innovation within the banking sector.

Authors' Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

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

Ethical Considerations

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

References

[1] A. Kanas, D. Vasiliou, and N. Eriotis, "Credit risk and profitability in banking: Evidence from international banks," *Research in International Business and Finance,* vol. 72, p. 102260, 2025, doi: 10.1016/j.ribaf.2024.102260.

[2] V. Mahadevan, S. Subramaniam, and V. Srivastava, "Carbon Concentration in Bank Portfolios and Efficiency: The Role of Credit Risk and Capitalization," *Journal of International Financial Markets, Institutions and Money,* vol. 109, p. 102327, 2026, doi: 10.1016/j.intfin.2026.102327.

[3] S. Sadegh Khani and F. Ramazani, "Credit risk and ranking of Bank Saderat Ilam customers using econometric methods and logistic regression model," in *Eighth National Conference on Economics, Management, and Accounting*, Shirvan, 2021. [Online]. Available: https://en.civilica.com/doc/1234991/.

[4] J. Maghsoodi, *Credit Scoring of Customers at Post Bank of Iran*. Tehran: Hanoon, 2021.

[5] A. Aghaei and F. Mohammadi, "Providing a credit risk prediction model for Bank Sepah customers using metaheuristic algorithms," *Banking Management Quarterly,* vol. 12, no. 46, pp. 67-84, 2022.

[6] A. Ghasemi Armaki, M. F. Fallah, and M. Alborzi, "Development and Explanation of a Customer Credit Scoring System for Banks Based on Hybrid Machine Learning Models: A Case Study of Bank Mellat," *Financial Management Perspective,* vol. 12, no. 37, pp. 69-94, 2022, doi: 10.52547/JFMP.12.37.69.

[7] M. Jalili, M. Khodaei Valehzaghar, and M. Kaneshlou, "Credit Scoring of Real Customers in the Country's Banking System," *Quarterly Journal of Quantitative Studies in*

Management, vol. 1, no. 3, pp. 127-148, 2023 2023. [Online]. Available: https://sid.ir/paper/191947/en.

[8] J. H. Plikas and D. Kenourgios, "Renewable Energy and Bank Credit Risk: The Mediating Role of Economic Growth," *Economic Modelling,* p. 107576, 2026, doi: 10.1016/j.econmod.2026.107576.

[9] V. Pourshahabi, H. Khodadadi Didani, and A. Taadolkhah, "Designing a Strategic Alliace Implementation Model in Customer-Based Organizations by Exploratory Mixed Method (Case study: Saman Bank)," *Strategic Management Researches,* vol. 27, no. 83, pp. 41-66, 2022. [Online]. Available: https://smr.journals.iau.ir/article_690906_2ec2f7d0a4fb6ed590b6cc50e26ec6c7.pdf.

[10] X. Yan, Y. Li, F. Nie, and R. Li, "Bank Customer Segmentation and Marketing Strategies Based on Improved DBSCAN Algorithm," *Applied Sciences,* vol. 15, no. 6, 2025.

[11] L. Yang and Y. Zhou, "Physical environment and its impact on customer perceptions of service quality in retail banking," *International Journal of Bank Marketing,* vol. 43, no. 2, pp. 210-228, 2025.

[12] M. J. Taghipourian, "The role of emotional branding on brand equity in the banking sector: Is customer experience a mediator?," *Dynamic Management and Business Analysis,* pp. 1-17, 2026. [Online]. Available: https://www.dmbaj.com/index.php/dmba/article/view/247.

[13] M. Jamshidi, "Determining and Ranking the Credit Scoring Indicators of Sepah Bank Customers," *Quarterly Journal of Islamic Financial and Banking Studies,* vol. 7, no. 17, pp. 47-66, 2022 2022. [Online]. Available: https://jifb.ibi.ac.ir/article_158635.html?lang=en.

[14] S. Abolghasemi, A. Nabavi Chashmi, E. Memarian, and M. Taghipour, "Explaining the Operational and Credit Risk Model in Banks with Emphasis on Ethical Validation of Customers," *Ethics in Science and Technology,* vol. 19, no. 2, pp. 185-190, 2024.

[15] A. Yāvarī, H. Jabbārī, and H. Panāhiyān, "Designing a Credit Risk Management Model with a Pathological Approach to Guarantees and Collaterals of Bank Facilities," *Technology in Entrepreneurship and Strategic Management Quarterly,* vol. 4, no. 2, pp. 1-22, 2025, doi: 10.61838/kman.jtesm.4.2.11.

[16] L. Abdollahzadeh Ramhormozi, A. H. Azh, and M. Khaki, "Evaluation and ranking service quality in the banking system and its impact on CRM Rishi Kant from the model (Case study: Melli Bank of Iran)," (in en), *Transactions on Data Analysis in Social Science,* vol. 5, no. 1, pp. 11-24, 2023, doi: 10.47176/TDASS/2023.11.