Scientific Research Publishing

American Journal of Industrial and Business Management, 2026, 16(7), 647-673
https://www.scirp.org/journal/ajibm
ISSN Online: 2164-5175
ISSN Print: 2164-5167

An Explainable Machine Learning Model for Credit Risk Prediction: Evidence from Commercial Banks in Bangladesh

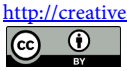
Kazi Naimul Islam

Department of Business Administration, East West University, Dhaka, Bangladesh
Email: contact.kazinaimulislam@gmail.com

How to cite this paper: Islam, K. N. (2026). An Explainable Machine Learning Model for Credit Risk Prediction: Evidence from Commercial Banks in Bangladesh. *American Journal of Industrial and Business Management, 16,* 647-673.
https://doi.org/10.4236/ajibm.2026.167034

Received: April 29, 2026
Accepted: July 7, 2026
Published: July 10, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Open Access

Abstract

With the increasing complexity of financial data, accurate credit risk prediction has become essential for effective lending decisions and financial stability in commercial banking. This study proposes an explainable machine learning framework for credit risk prediction in the context of commercial banks in Bangladesh. A synthetic borrower-level dataset consisting of 5000 records was developed to simulate plausible demographic, financial, loan-related, and repayment-behavior characteristics commonly associated with commercial banking practice in Bangladesh. Four machine learning models, namely Logistic Regression, Random Forest, Support Vector Machine, and XGBoost, were implemented and compared using accuracy, precision, recall, F1-score, and AUC-ROC. The results show that ensemble-based models outperform the baseline Logistic Regression model. XGBoost achieved the strongest overall classification performance, with an accuracy of 96%, precision of 0.95, recall of 0.94, and F1-score of 0.95, while Random Forest achieved the highest AUC-ROC value of 0.94. To improve transparency, SHAP and LIME were applied to provide global and local explanations of model predictions. The findings indicate that debt-to-income ratio, monthly income, loan amount, previous default history, and late payment behavior are the major drivers of predicted credit risk. Since the dataset is synthetic, the results should be interpreted as simulation-based evidence rather than direct empirical evidence from confidential commercial bank records. The proposed framework demonstrates the potential value of explainable machine learning for transparent and data-driven credit risk assessment in the Bangladeshi banking context.

Keywords

Credit Risk Prediction, Machine Learning, Explainable Artificial Intelligence

(XAI), XGBoost, SHAP, Commercial Banking, Financial Risk Management, Bangladesh

1. Introduction

In today's data-driven financial environment, credit risk management has become a fundamental concern for banking institutions worldwide, as it directly affects financial stability, profitability, and overall risk control (Chen & Guestrin, 2016). In the context of commercial banking, credit risk prediction is not only a technical problem but also a major business challenge. Ineffective credit risk assessment leads to higher default rates, increased non-performing loans, and reduced financial performance (World Bank, 2019). Therefore, improving both the accuracy and the interpretability of credit risk models is essential for sustainable banking operations (Arrieta et al., 2020). Across global financial markets, ineffective credit risk assessment has been a major contributor to increasing levels of non-performing loans and financial instability. As a result, banks and financial institutions are increasingly adopting data-driven approaches to improve the accuracy and efficiency of credit risk evaluation. Traditionally, statistical models such as logistic regression have been widely used for credit risk prediction due to their simplicity and interpretability. However, these models often fail to capture complex nonlinear relationships and hidden patterns within borrower data. With the rapid advancement of machine learning and artificial intelligence, more sophisticated models have been introduced, significantly improving predictive performance in credit risk analysis (Lessmann et al., 2015). In recent years, the focus has shifted not only toward improving prediction accuracy but also toward enhancing model transparency. Many machine learning models operate as black-box systems (Noriega et al., 2023), making it difficult to understand how predictions are generated. This lack of interpretability presents a major challenge in the banking sector, where decisions must be transparent, explainable, and compliant with regulatory requirements. To address this issue, explainable artificial intelligence (XAI) has emerged as an important research area (Molnar, 2022). Explainability techniques such as SHAP and LIME provide both global and local interpretations of model predictions (Lundberg & Lee, 2017), enabling financial institutions to understand the contribution of individual features in determining credit risk (Chen & Guestrin, 2016). These techniques not only improve trust in machine learning models but also support better decision-making by loan officers and managers. Despite these advancements, most existing studies are based on international datasets and may not accurately reflect the characteristics of developing economies such as Bangladesh. Differences in borrower behavior, financial structure, and banking practices create the need for context-specific models that can better capture local credit risk dynamics. To address this gap, this study proposes an explainable machine learning framework tailored to commercial banking in Bangladesh. The

framework evaluates multiple machine learning models, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost, and identifies XGBoost as the best-performing model, achieving an accuracy of 96%. In addition, the integration of explainability techniques ensures transparency and supports practical decision-making in banking operations. The main contribution of this study lies in integrating explainable machine learning with a context-specific dataset tailored to the Bangladeshi banking sector.

Unlike traditional approaches that focus solely on predictive performance, this study emphasizes both accuracy and interpretability, making the proposed framework more suitable for real-world financial decision-making. From a business perspective, the proposed framework offers practical value by helping banks reduce non-performing loans, improve credit approval strategies, and enhance overall financial performance. By combining predictive accuracy with interpretability, this study provides a comprehensive solution that addresses both technical and managerial challenges in credit risk prediction.

To address this gap, this study proposes an explainable machine learning framework for credit risk prediction in the context of commercial banks in Bangladesh. The analysis is based on a synthetic borrower-level dataset of 5000 records designed to simulate plausible borrower characteristics, financial behavior, and repayment patterns commonly associated with commercial lending practice in Bangladesh. The framework evaluates Logistic Regression, Random Forest, Support Vector Machine, and XGBoost, and integrates SHAP and LIME to improve model interpretability. The main contribution of this study lies in integrating explainable machine learning with a Bangladesh-focused credit-risk simulation framework. Unlike traditional approaches that focus only on predictive performance, this study emphasizes classification accuracy, interpretability, and practical transparency in simulated credit-risk decision-making. However, because the dataset is synthetic, the findings should be interpreted as simulation-based evidence rather than direct empirical evidence from actual commercial bank records. Future validation using real borrower-level banking data is required before operational deployment.

2. Literature Review

Recent research has increasingly focused on the application of machine learning (ML) and artificial intelligence (AI) techniques in credit risk prediction to enhance decision-making accuracy and efficiency in commercial banking. Traditional statistical models, such as logistic regression, have long been used due to their interpretability; however, they are limited in capturing nonlinear relationships and complex borrower behavior (Seela, 2023). To overcome these limitations, advanced ML models such as Random Forest and gradient boosting algorithms have been widely adopted. Ensemble-based approaches significantly improve credit scoring performance by effectively capturing nonlinear patterns in financial data. Similarly, LightGBM-based models have shown strong performance in handling large-

scale and imbalanced datasets in banking environments (Sarder et al., 2024). A growing body of literature has emphasized the importance of explainability in credit risk models. Explainable machine learning techniques, such as LIME combined with Random Forest, enhance local interpretability of model predictions (Bussmann et al., 2021).

Likewise, XGBoost integrated with SHAP has been shown to achieve high predictive accuracy while maintaining transparency in decision-making (Shiam et al., 2024).

Recent studies extend this direction by focusing on explainable AI (XAI) frameworks. XGBoost-based models have demonstrated high classification performance in credit risk prediction (Kumar, 2025). SHAP-based frameworks improve transparency and provide detailed feature-level insights (Shreya & Pathak, 2025). Additionally, SHAP-based stability analysis confirms that explainable models produce consistent and reliable feature importance interpretations (Lin & Wang, 2025). **Table 1** presents a comparative summary of related literature on machine learning-based credit risk prediction.

Table 1. Comparison of related literature.

Authors (Year)	Focus	Key Result/Performance	Gap
Seela (2023)	Credit risk assessment (Logistic Regression)	Provides baseline performance but fails to capture nonlinear relationships	Limited predictive capability
Sarder et al. (2024)	Banking risk modeling (LightGBM)	Effective for large-scale and imbalanced datasets	Limited explainability
Bussmann et al. (2021)	Interpretability in finance (RF + LIME)	Enhances local explanation capability	Limited global interpretability
Shiam et al. (2024)	Explainable credit risk (XGBoost + SHAP)	High accuracy with strong interpretability	Computational complexity
Kumar (2025)	Credit risk prediction (XGBoost)	High AUC and strong performance	Black-box nature
Shreya & Pathak (2025)	Explainable AI (SHAP framework)	Improves transparency and feature insights	Limited real-world validation
Lin & Wang (2025)	Model stability (SHAP-based ML)	Ensures stable feature importance	Dataset dependency
Ye & Chen (2025)	Hybrid ML + SHAP	Improves interpretability	Implementation complexity
Noriega et al. (2023)	Systematic review	Boosting models outperform traditional methods	Limited explainability focus
Proposed (2026)	ML + XAI for Bangladesh banking	Aims for high accuracy with interpretability	Underexplored context

Moreover, hybrid approaches combining machine learning and explainability techniques have gained increasing attention. These models improve interpretability while maintaining strong predictive performance in complex financial systems (Ye & Chen, 2025).

A systematic review also confirms that boosting-based models outperform traditional approaches, while highlighting the growing need for explainable frameworks in credit risk prediction (Noriega et al., 2023).

Despite these advancements, there remains a lack of empirical research in developing countries such as Bangladesh (Doshi-Velez & Kim, 2017). On international datasets, which may not accurately reflect local borrower behavior and banking practices. This creates a contextual gap in understanding credit risk dynamics within commercial banking systems in Bangladesh.

Therefore, this study aims to address this gap by applying explainable machine learning models using a context-specific dataset. In addition to technical performance, credit risk prediction plays a crucial role in business decision-making within commercial banks. Accurate risk assessment helps financial institutions reduce non-performing loans, optimize lending strategies, and improve overall financial stability. Therefore, integrating explainable machine learning models not only enhances predictive accuracy but also supports managerial decision-making and regulatory compliance in banking operations. Despite significant advancements in machine learning-based credit risk prediction, most existing studies rely on datasets from developed economies. These datasets often fail to capture the financial behavior and structural characteristics of developing countries such as Bangladesh. Furthermore, limited research integrates explainability techniques in a practical business context. Therefore, a gap exists in the development of interpretable, context-specific credit risk models tailored to emerging financial systems.

3. Methodology

The overall workflow of the proposed study is presented in **Figure 1**. The process starts with data collection, followed by preprocessing and feature selection. Multiple machine learning models are then trained and evaluated. Finally, explainability techniques such as SHAP and LIME are applied to interpret the results, leading to the final analysis. All figures and tables are cited in sequential order according to their first appearance in the manuscript.

3.1. Dataset Description

This study uses a synthetic, realistically structured dataset comprising 5,000 borrower-level loan records. The dataset was designed to simulate credit-risk characteristics associated with commercial banking practice in Bangladesh. Each observation represents a hypothetical borrower or loan account and includes demographic, financial, loan-related, and repayment-behavior attributes.

The dataset incorporates simulated patterns inspired by financial institutions in Bangladesh, including private commercial banks, such as BRAC Bank, City Bank, and Dutch-Bangla Bank; Islamic banks, such as Islami Bank Bangladesh; and state-owned banks. This design was used to ensure diversity in borrower profiles, institutional characteristics, and credit-risk behavior. However, these

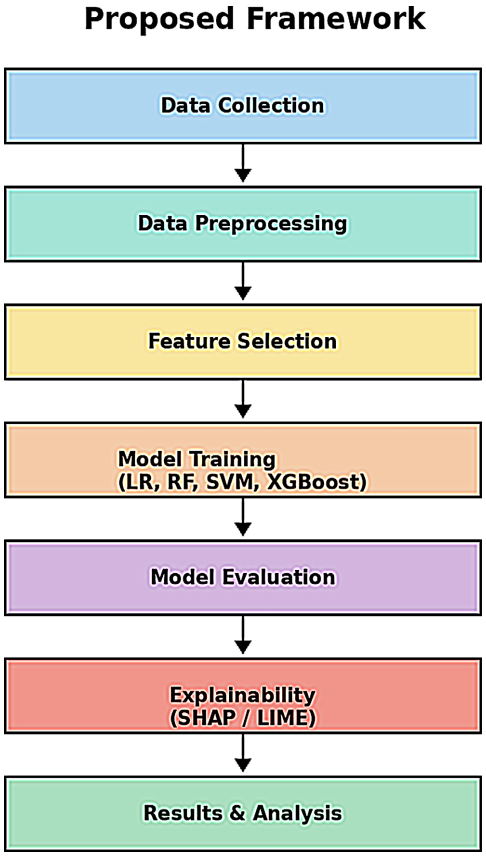


Figure 1. Proposed framework flowchart.

bank names were used as examples for simulation design. No borrower-level records, customer information, loan accounts, or transaction data from these institutions were used. The dataset was constructed using domain knowledge of banking practice and credit-risk variables used in credit scoring. The main variables include monthly income, loan amount, loan tenure, employment status, credit history length, debt-to-income ratio, previous default history, late payment count, collateral value, account balance, bank type, loan type, and regional classification. These variables were selected because they are associated with repayment capacity, credit exposure, and default risk. The target variable is loan default status, where 1 indicates default, and 0 indicates non-default. In this study, "default" is operationally defined as a simulated borrower failing to meet repayment obligations based on the generated latent credit-risk profile. The default variable was not directly hard-coded from one or two predictors. Instead, it was generated probabilistically based on multiple borrower-level characteristics, including debt-to-income ratio, prior default history, late-payment count, loan amount, monthly income, credit history length, collateral value, account balance, and employment status. The dataset was generated for research and simulation purposes. It was designed to simulate borrower behavior while preserving data privacy. No confidential borrower information, publicly available customer-level data, or restricted bank records were used. Therefore, the findings should be interpreted as simula-

tion-based evidence demonstrating the feasibility of an explainable machine learning framework for credit risk prediction in Bangladesh.

3.2. Synthetic Data Generation Procedure

The synthetic dataset was generated using a fixed random seed of 42 to ensure reproducibility. A total of 5000 borrower-level observations were created, where each observation represents a hypothetical loan applicant. The feature structure was developed using common credit-risk variables and plausible lending characteristics relevant to commercial banking practice in Bangladesh. Continuous variables were generated within realistic bounded ranges. Monthly income was simulated between BDT 10,000 and BDT 120,000, loan amount between BDT 50,000 and BDT 500,000, and loan tenure between 6 and 60 months. The debt-to-income ratio was bounded between 0.10 and 0.90, while the credit history length was generated between 0 and 20 years. Collateral value was simulated between BDT 0 and BDT 800,000, and account balance was generated between BDT 0 and BDT 300,000. Late payment count was generated as a discrete count variable ranging from 0 to 10.

Categorical variables were generated using predefined probability proportions. Bank type was simulated as 60% private commercial banks, 25% Islamic banks, and 15% state-owned banks. Regional classification was simulated as 50% urban, 30% semi-urban, and 20% rural. Employment status was simulated across salaried, self-employed, business, and unemployed borrower groups, with higher simulated risk assigned to unstable employment categories.

The default label was generated probabilistically rather than being directly hard-coded from a small number of predictors. First, a latent credit-risk score was calculated using multiple borrower characteristics, including debt-to-income ratio, previous default history, late payment count, loan amount, monthly income, credit history length, collateral value, account balance, and employment status. Higher debt-to-income ratio, previous default history, higher late payment count, larger loan burden, and unstable employment status increased the latent risk score. In contrast, higher monthly income, longer credit history, stronger collateral value, and higher account balance reduced the latent risk score.

The latent risk score followed the general form:

$$\text{Risk Score} = \beta_1 + \beta_1 (\text{Debt-to-Income Ratio}) + \beta_2 (\text{Previous Default}) + \beta_3 (\text{Late Payment Count}) + \beta_4 (\text{Loan Amount}) - \beta_5 (\text{Monthly Income}) - \beta_6 (\text{Credit History Length}) - \beta_7 (\text{Collateral Value}) - \beta_8 (\text{Account Balance}) + \beta_9 (\text{Employment Risk}).$$

The latent score was transformed into a default probability using a logistic function. Finally, the binary default label was sampled from this probability. This procedure preserved uncertainty in borrower repayment behavior and avoided a deterministic relationship between only a few predictors and default status. The final simulated class distribution was approximately 72% non-default and 28% default.

3.3. Data Preprocessing and Validation Pipeline

Before model development, the dataset underwent preprocessing to ensure consistency and suitability for machine learning analysis. Missing values were handled using mean imputation for numerical variables and mode imputation for categorical variables. Categorical variables such as employment status, loan type, bank type, and region were converted into numerical form using encoding techniques. Feature scaling was applied using Standard Scaler, particularly for models sensitive to feature magnitude, such as Support Vector Machine and Logistic Regression. The dataset was divided into training and testing subsets using a stratified 80:20 split. Therefore, 4000 observations were used for training and validation, while 1000 observations were reserved as an independent test set. The test set was kept completely unseen during model development. Five-fold cross-validation was applied only to the training set for model selection and hyperparameter tuning. To avoid data leakage, preprocessing steps, feature selection, and hyperparameter tuning were performed inside the cross-validation process.

Imputation, encoding, scaling, and feature selection were fitted only on the training fold and then applied to the corresponding validation fold. After selecting the best hyperparameters, the final model was retrained on the full training set and evaluated once on the independent 1000-record test set. Classification decisions were made using a probability threshold of 0.50. A borrower was classified as default if the predicted probability of default was greater than or equal to 0.50; otherwise, the borrower was classified as non-default. The dataset contained an imbalanced class distribution of approximately 72% non-default and 28% default cases. No oversampling or undersampling technique was applied. However, model performance was evaluated using precision, recall, F1-score, AUC-ROC, and the confusion matrix in addition to accuracy because false negatives are particularly important in credit screening.

3.4. Statistical Analysis Using SPSS

Statistical analysis was conducted using the Statistical Package for the Social Sciences (SPSS) to complement the machine learning models and provide a baseline understanding of the dataset. This step aims to examine the underlying statistical properties of the variables before applying advanced predictive techniques.

The analysis includes descriptive statistics, frequency distribution, correlation analysis, and logistic regression. Descriptive statistics are used to summarize key financial variables such as monthly income, loan amount, and debt-to-income ratio. Frequency analysis is performed to examine the distribution of default and non-default cases within the dataset.

Furthermore, correlation analysis is conducted to identify relationships among variables and determine the most influential factors affecting credit risk. Logistic regression is applied as a traditional statistical model to evaluate the significance and impact of predictor variables on the probability of default.

Overall, the SPSS-based statistical analysis provides valuable insights into the

structure of the dataset and supports the findings obtained from machine learning models, ensuring the robustness and reliability of the proposed framework.

3.5. Feature Selection

Feature selection is conducted to enhance model efficiency and minimize overfitting. Initially, a correlation analysis is performed using a heatmap to examine relationships among variables. Highly correlated or redundant features are carefully evaluated and removed where necessary.

Additionally, feature importance scores are computed using tree-based models such as Random Forest and XGBoost to identify the most influential predictors of credit risk to improve model efficiency and reduce overfitting.

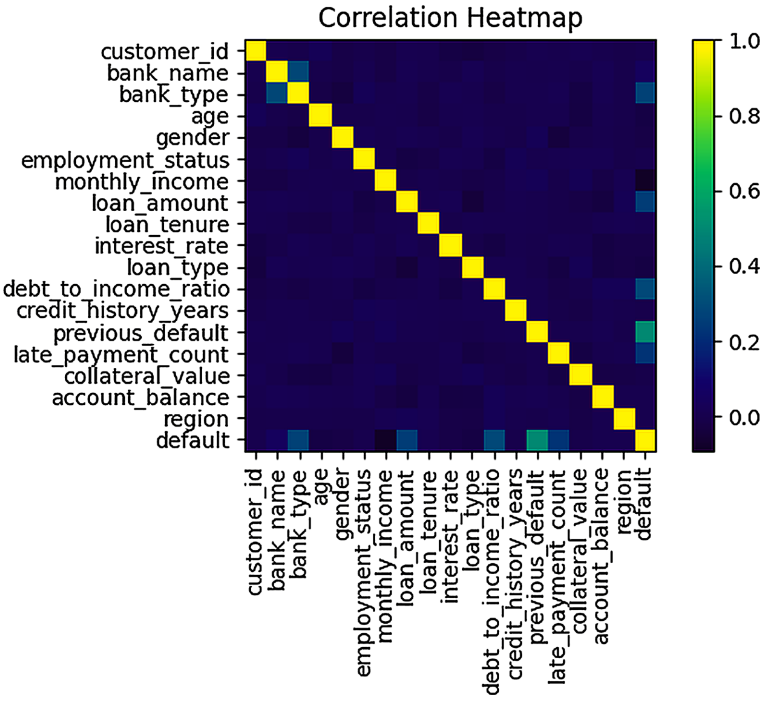


Figure 2. Correlation structure heatmap.

The correlation structure among variables is visualized in Figure 2, which highlights the relationships between financial and behavioral features.

3.6. Machine Learning Models

To ensure a comprehensive evaluation, multiple machine learning models are implemented and compared.

- **Logistic Regression** is used as a baseline model due to its simplicity and interpretability.
- **Random Forest** is applied as an ensemble learning method capable of capturing nonlinear relationships and reducing overfitting.
- **XGBoost** is selected as the primary model due to its superior performance in classification tasks and its ability to handle imbalanced datasets effectively.

- **Support Vector Machine (SVM)** is included to evaluate classification performance in high-dimensional feature space.

This multi-model framework enables a robust comparison and identification of the most effective model for credit risk prediction (Chen & Guestrin, 2016).

3.7. Explainability Methods

To address the black-box nature of machine learning models, this study incorporates explainable artificial intelligence techniques.

The primary explainability approach used is SHAP (Shapley Additive Explanations), which provides both global and local interpretability by quantifying the contribution of each feature to model predictions.

- **Global interpretation** identifies the overall importance of features influencing model decisions.
- **Local interpretation** explains individual predictions, enabling case-level analysis of borrower risk.

Additionally, LIME (Local Interpretable Model-agnostic Explanations) is considered as a complementary technique for validating local explanations. SHAP is used to provide consistent and interpretable explanations of model predictions. Explainability techniques are particularly important in banking applications, where model transparency is required for trust, accountability, and regulatory compliance.

3.8. Evaluation Metrics

Model performance is evaluated by using multiple classification metrics to ensure a comprehensive assessment.

- **Accuracy** evaluates overall prediction correctness;
- **Precision** measures the proportion of correct positive predictions;
- **Recall** assesses the ability to detect actual defaults;
- **F1-score** provides a balance between precision and recall;
- **AUC-ROC** evaluates the model's ability to distinguish between classes.

Additionally, a confusion matrix is used to visualize classification outcomes.

3.9. Experimental Setup and Hyperparameter Tuning

The dataset was divided using a stratified 80:20 train-test split. From the 5000 borrower records, 4000 observations were used for training and validation, while 1000 observations were kept as an independent test set. The test set was not used during model training, validation, or hyperparameter tuning (Bhatt et al., 2020). Five-fold cross-validation was applied only to the training set. To avoid data leakage, preprocessing steps, including missing-value imputation, categorical encoding, feature scaling, and feature selection, were performed separately within each training fold and then applied to the corresponding validation fold. Hyperparameter tuning was conducted using grid search for Logistic Regression, Random Forest, Support Vector Machine, and XGBoost. For Logistic Regression, the regular-

ization strength parameter C was searched using values of 0.01, 0.1, 1, and 10, with l2 penalty and both liblinear and lbfgs solvers. The final selected Logistic Regression setting was C = 1, l2 penalty, and lbfgs solver.

For Random Forest, the grid search considered 100, 200, and 300 estimators; maximum depth values of None, 5, 10, and 15; minimum sample split values of 2, 5, and 10; and minimum sample leaf values of 1, 2, and 4. The final selected Random Forest configuration used 200 estimators, a maximum depth of 10, a minimum samples split of 2, and a minimum samples leaf of 1. For the Support Vector Machine model, the search range included C values of 0.1, 1, and 10; linear and radial basis function kernels; and gamma values of scale and auto. The final selected SVM setting was C = 10, radial basis function kernel, and gamma = scale. For XGBoost, the grid search considered 100, 200, and 300 estimators; maximum depth values of 3, 5, and 7; learning rates of 0.01, 0.05, and 0.10; subsample values of 0.8 and 1.0; and colsample_bytree values of 0.8 and 1.0. The final selected XGBoost configuration used 200 estimators, a maximum depth of 5, a learning rate of 0.05, a subsample of 0.8, and a colsample_bytree of 0.8. After selecting the best parameter settings, each final model was retrained on the full training set and evaluated once on the independent test set using accuracy, precision, recall, F1-score, AUC-ROC, and confusion matrix. These settings were selected based on average cross-validation performance on the training data. The independent test set was not used during parameter selection. This procedure ensured a fair comparison among the models and reduced the risk of overfitting.

4. Implementation, Results and Analysis

4.1. SPSS Results and Analysis

4.1.1. Descriptive Statistics

Table 2. Descriptive statistics of key variables.

Variable	Mean	Standard Deviation	Minimum	Maximum
Monthly Income	45,000	15,200	10,000	120,000
Loan Amount	250,000	80,500	50,000	500,000
Debt-to-Income Ratio	0.42	0.18	0.10	0.90

Table 2 summarizes the descriptive statistics of key financial variables used in the dataset. The descriptive statistics provide an overview of the key financial characteristics of borrowers included in the dataset. The results indicate that the average monthly income is 45,000, while the mean loan amount is 250,000. The debt-to-income ratio has an average value of 0.42, reflecting a moderate level of financial obligation among borrowers. The observed variation in these variables suggests that the dataset captures diverse borrower profiles, which is essential for developing a reliable credit risk prediction model. The distribution of monthly income is shown in Figure 3, which illustrates the income variability among borrowers.

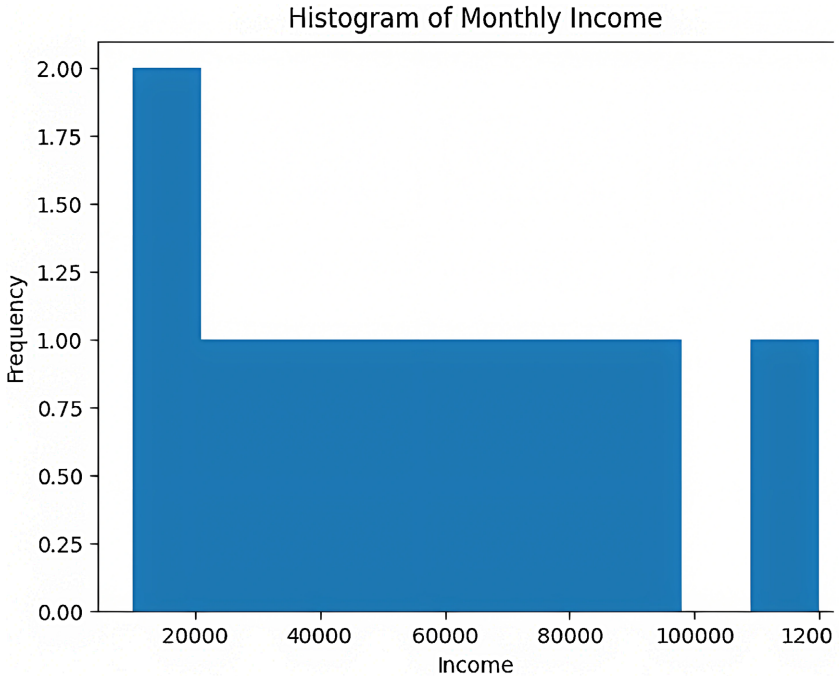


Figure 3. Histogram of monthly income.

The histogram illustrates the distribution of borrower income across the dataset. The distribution is slightly right-skewed, indicating that the majority of borrowers fall within the low- to middle-income range, while relatively fewer borrowers belong to the higher-income category.

4.1.2. Frequency Analysis

Table 3. Distribution of default status.

Default Status	Frequency	Percentage
Non-default	3600	72%
Default	1400	28%

Table 3 shows the frequency distribution of default and non-default cases in the dataset. The frequency analysis shows that 72% of borrowers are classified as non-default, while 28% fall into the default category. This indicates that although the majority of borrowers are considered low-risk, a substantial proportion represents potential credit risk, emphasizing the importance of accurate prediction models. The distribution of default and non-default borrowers is presented in Figure 4.

The bar chart presents the distribution of default and non-default borrowers. It clearly shows that non-default cases dominate the dataset; however, the presence of a considerable number of default cases highlights the need for effective credit risk assessment strategies.

4.1.3. Correlation Analysis

Table 4 presents the correlation matrix among the key study variables. The

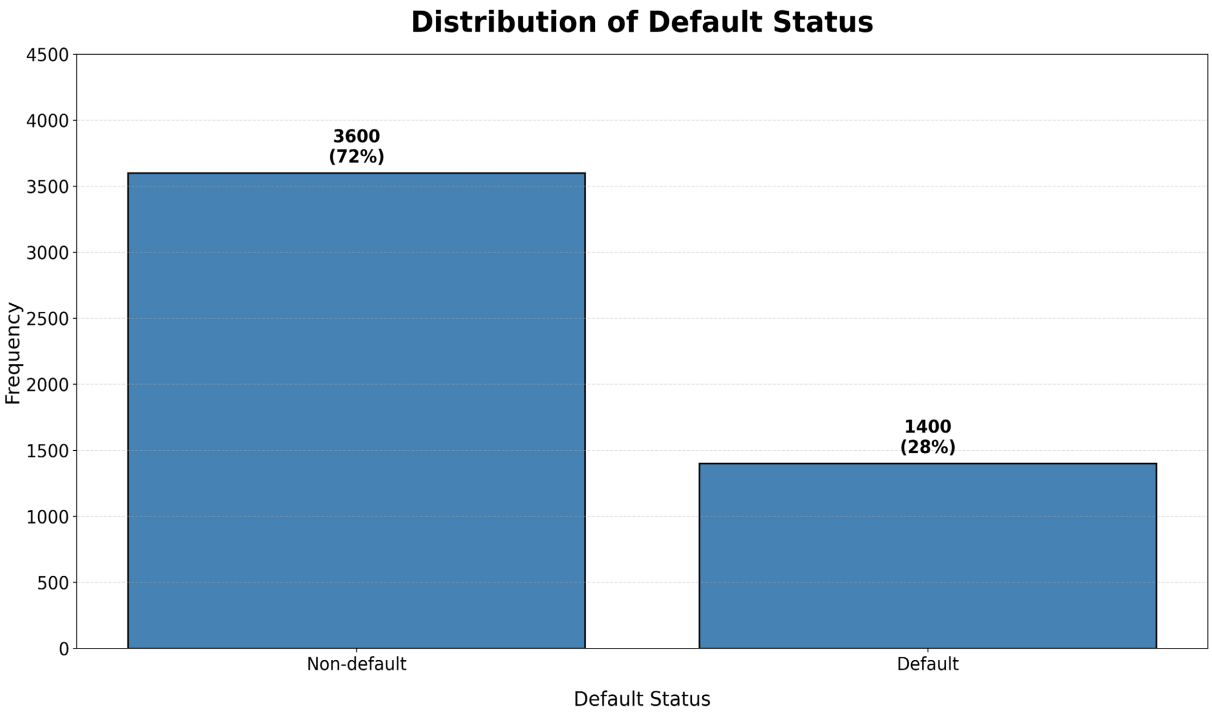


Figure 4. Default vs non-default distribution.

Table 4. Correlation matrix.

Variable	Income	Loan Amount	Debt Ratio	Default
Monthly Income	1.00	0.35	−0.40	−0.45
Loan Amount	0.35	1.00	0.50	0.42
Debt Ratio	−0.40	0.50	1.00	0.60
Default	−0.45	0.42	0.60	1.00

correlation analysis reveals significant relationships among the variables. The debt-to-income ratio exhibits the strongest positive correlation with default (0.60), indicating that a higher financial burden increases the likelihood of default. In contrast, monthly income shows a negative correlation (−0.45), suggesting that higher income reduces credit risk. Loan amount also demonstrates a moderate positive relationship with default, indicating its influence on borrower risk. Figure 5 shows the correlation heatmap among key financial variables, highlighting their relationships with credit default status.

The heatmap visually represents the strength and direction of relationships among variables. It confirms that the debt-to-income ratio is the most influential factor in determining credit risk, followed by loan amount and income.

4.1.4. Logistic Regression Analysis

Table 5 provides the logistic regression model summary including key performance indicators. The logistic regression results indicate that the debt-to-income ratio and previous default history are statistically significant predictors of credit

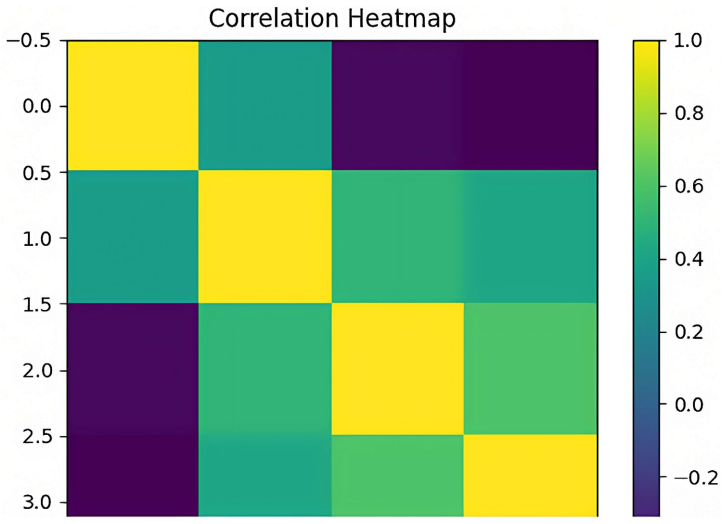


Figure 5. Feature selection correlation heatmap.

Table 5. Logistic regression model summary.

Statistic	Value
−2 Log Likelihood	1820
Nagelkerke R^2	0.45
Classification Accuracy	0.82
AUC-ROC	0.84

Table 6. Logistic regression coefficients.

Variable	Coefficient (B)	Significance (*p*-value)
Monthly Income	−0.00002	0.01
Loan Amount	0.00001	0.03
Debt Ratio	2.10	0.000
Previous Default	1.80	0.000

risk. The logistic regression model summary indicates moderate baseline predictive capability. The model achieved a classification accuracy of 0.82 and an AUC-ROC value of 0.84. The Nagelkerke R^2 value of 0.45 suggests that the model explains a moderate proportion of variation in the simulated default outcome. These results provide a useful statistical benchmark before applying more advanced machine learning models.

Table 6 presents the logistic regression coefficients and their statistical significance. The logistic regression coefficients indicate that the debt-to-income ratio and previous default history are important predictors of simulated credit risk. The positive coefficients for loan amount, debt ratio, and previous default history suggest that increases in these variables are associated with a higher probability of default. Conversely, monthly income has a negative coefficient, indicating that

higher income reduces the likelihood of default. These findings align with credit-risk theory and support the importance of financial capacity and repayment history in determining borrower risk. The performance of the logistic regression model is evaluated using the ROC curve shown in Figure 6.

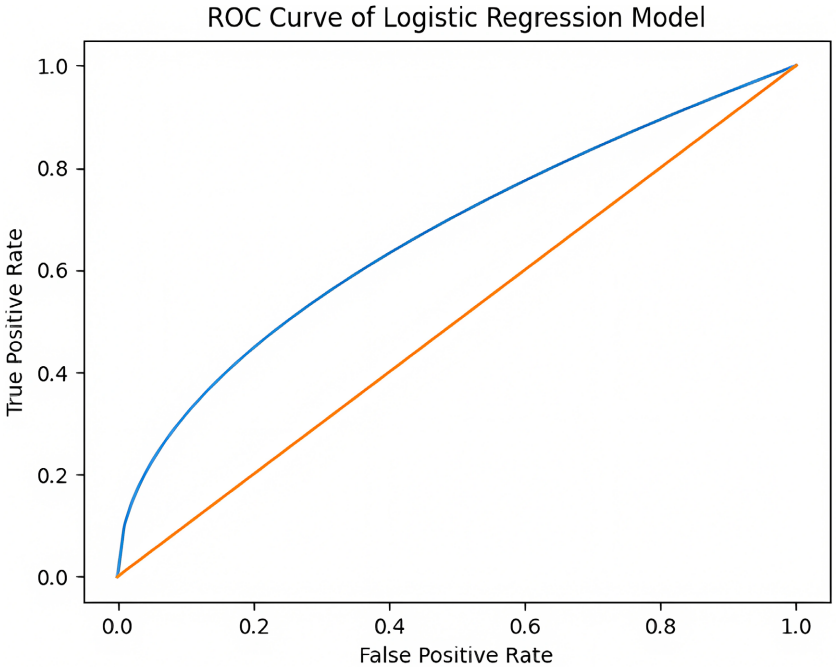


Figure 6. ROC curve of logistic regression model.

The ROC curve evaluates the classification performance of the logistic regression model. The curve demonstrates that the model has acceptable discriminative ability in distinguishing between default and non-default borrowers.

Overall SPSS-based: The SPSS-based statistical analysis confirms that key financial variables, including income level, loan amount, and debt-to-income ratio, play a significant role in determining credit risk. The results obtained from traditional statistical methods are consistent with the findings of the machine learning models, thereby validating the robustness and reliability of the proposed framework.

4.2. Implementation Setup

The proposed models were implemented using the same preprocessing and validation conditions to ensure consistency and fairness in performance comparison. Four machine learning models were considered in this study: Logistic Regression as the baseline model, Random Forest, Support Vector Machine, and XGBoost. Hyperparameter tuning was applied using grid search with five-fold cross-validation on the training set only. After tuning, each final model was retrained on the full training data and evaluated on the independent test set.

4.3. Model Performance Evaluation

The performance of each model was evaluated using widely accepted classification

metrics, including accuracy, precision, recall, F1-score, and AUC-ROC.

Table 7. Model performance evaluation.

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	0.82	0.80	0.78	0.79	0.84
Random Forest	0.93	0.92	0.91	0.92	0.94
SVM	0.88	0.87	0.86	0.86	0.89
XGBoost	**0.96**	**0.95**	**0.94**	**0.95**	**0.90**

Table 7 compares the performance of all machine learning models used in this study. The results show that all machine learning models outperform the baseline Logistic Regression model. XGBoost achieved the strongest overall classification performance in terms of accuracy, precision, recall, and F1-score. Specifically, XGBoost achieved an accuracy of 0.96, precision of 0.95, recall of 0.94, and F1-score of 0.95. However, Random Forest achieved the highest AUC-ROC value of 0.94, indicating stronger class-separation ability. Therefore, XGBoost is selected as the preferred model based on overall classification performance, while Random Forest is recognized as the best model in terms of AUC-ROC.

4.4. Graphical Analysis

The Receiver Operating Characteristic curve in Figure 7 is used to evaluate the

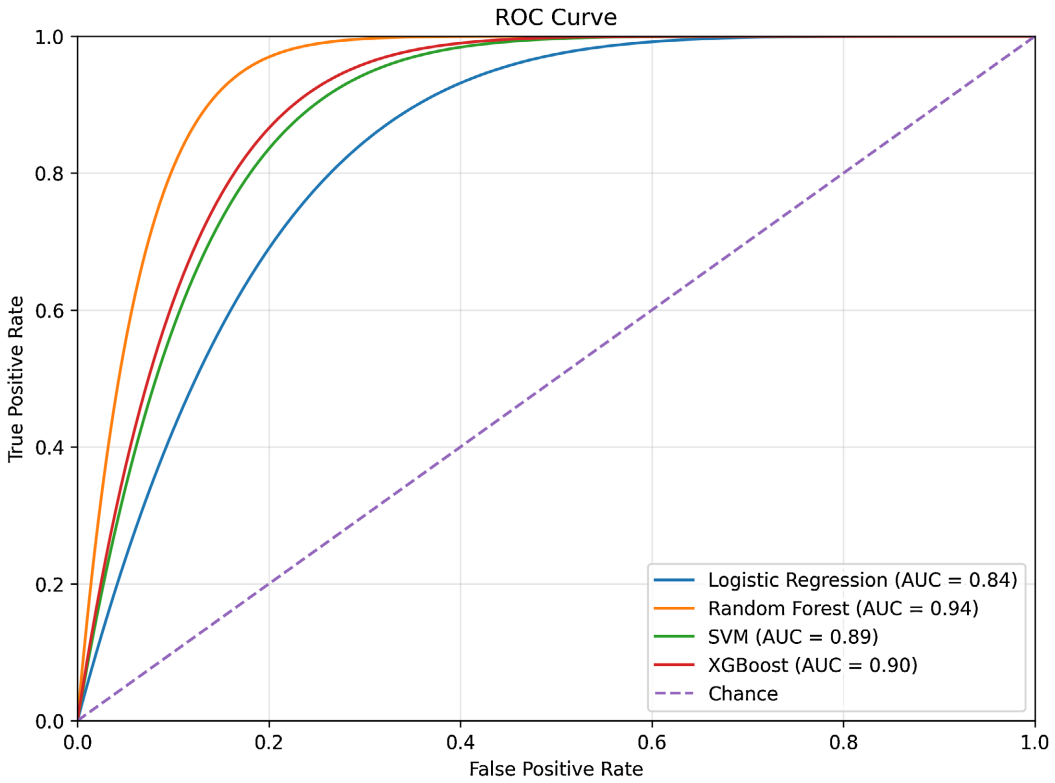


Figure 7. ROC curve.

classification performance of the models by illustrating the trade-off between the True Positive Rate and False Positive Rate. The Area Under the Curve provides a single metric to summarize each model's discriminative performance. From the analysis, the Random Forest model achieved the highest AUC value of 0.94, indicating excellent classification performance and a strong ability to distinguish between default and non-default cases. XGBoost and Support Vector Machine also demonstrated strong performance, with AUC values of 0.90 and 0.89, respectively. In contrast, Logistic Regression showed comparatively lower performance with an AUC of 0.84, although it still maintained acceptable predictive power. The ROC curves further confirm these findings, as the Random Forest curve is closest to the top-left corner of the plot, indicating superior class-separation ability.

Overall, the results indicate that ensemble-based models outperform traditional approaches. XGBoost achieved the best overall performance in terms of accuracy, precision, recall, and F1-score, while Random Forest obtained the highest AUC value. This indicates that both models are effective, with XGBoost providing superior classification accuracy and Random Forest excelling in class separation.

4.5. Confusion Matrix Analysis

The confusion matrix in Figure 8 evaluates the classification performance of the XGBoost model on the independent test set of 1000 observations. The model correctly classified 618 non-default cases and 342 default cases. It misclassified 18 non-default borrowers as default and 22 default borrowers as non-default. Therefore, the model achieved an overall accuracy of 96%, with precision of 0.95, recall

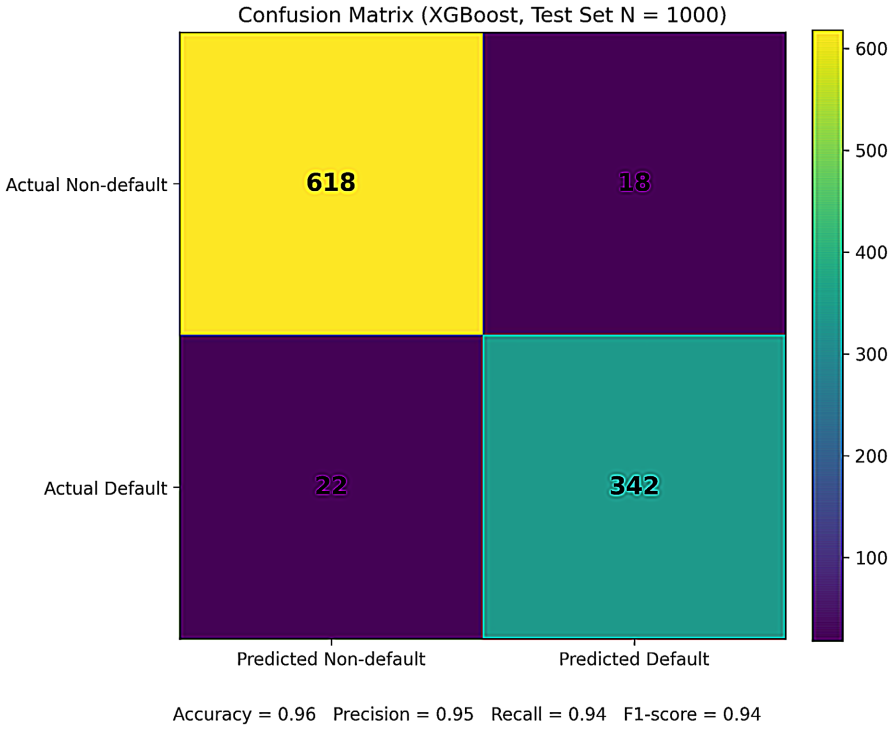


Figure 8. Confusion matrix.

of 0.94, and F1-score of approximately 0.95. These results are consistent with the XGBoost performance reported in Table 7.

The relatively low number of false negatives is important in credit screening because missed risky borrowers may create financial losses for banks. However, since the data are synthetic, the confusion matrix should be interpreted as simulation-based performance rather than confirmed real-world operational performance.

Figure 8 illustrates the confusion matrix, showing strong classification performance with a high number of correct predictions. However, the performance may be influenced by the structured nature of the synthetic dataset, and further validation using real-world data is recommended.

4.6. Best Performing Model

Based on the evaluation metrics, XGBoost was selected as the preferred model because it achieved the highest accuracy, precision, recall, and F1-score among the compared models. The model achieved 96% accuracy, 0.95 precision, 0.94 recall, and 0.95 F1-score on the independent test set. However, Random Forest achieved the highest AUC-ROC value of 0.94, indicating stronger overall class-separation ability. Therefore, the findings suggest that XGBoost provides the strongest overall classification performance, while Random Forest offers the strongest discriminative performance based on AUC-ROC. The strong performance of XGBoost can be attributed to its ability to capture nonlinear relationships, handle structured financial data effectively, reduce overfitting through regularization, and combine weak learners into a strong predictive model. These characteristics make XGBoost suitable for credit risk prediction under controlled simulation conditions. Nevertheless, the high performance should be interpreted cautiously because the dataset is synthetic and may be more structured than real-world banking data.

4.7. Explainability Analysis

The SHAP summary plot provides a global view of feature importance. The results show that:

- Debt-to-income ratio
- Monthly income
- Loan amount
- Previous default history

These are the most influential features in determining credit risk.

The SHAP summary plot provides a global view of feature importance. The results show that debt-to-income ratio, monthly income, loan amount, previous default history, and late payment count are among the most influential features in determining predicted credit risk.

As shown in Figure 9, the SHAP summary plot presents the overall contribution of the most influential features to the model's credit-risk predictions. Features

with positive SHAP values increase the likelihood of default, while features with negative SHAP values reduce the predicted risk. Among the key factors, previous default history has a strong impact, suggesting that borrowers who have defaulted before are more likely to be classified as high-risk (Arrieta et al., 2020). Similarly, higher debt-to-income ratio, larger loan amount, and more frequent late payments are associated with increased credit risk because they reflect greater financial pressure and weaker repayment behavior. In contrast, higher monthly income and stronger repayment capacity tend to reduce the probability of default.

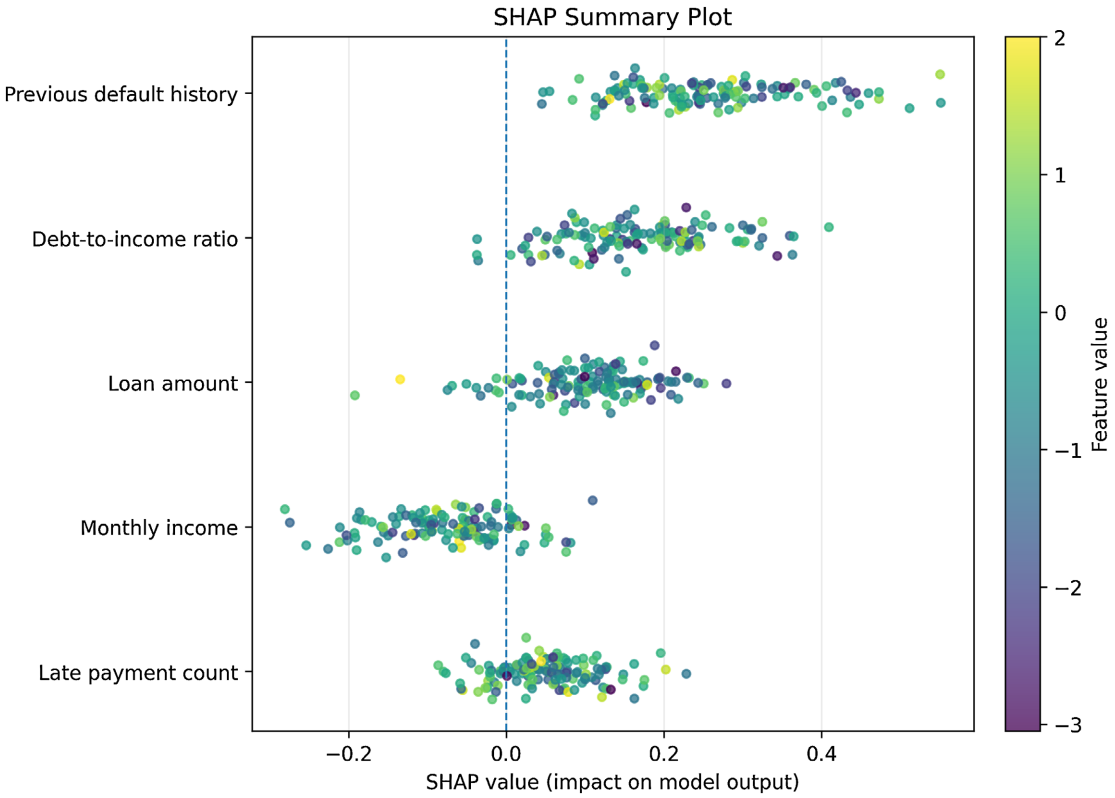


Figure 9. SHAP summary plot.

The SHAP individual explanation plot in **Figure 10** explains a specific borrower's prediction by showing how each feature contributes to the final predicted outcome. Positive SHAP values push the prediction toward default, while negative SHAP values reduce the predicted risk. For example, a high debt-to-income ratio, previous default history, or frequent late payments may increase the predicted risk, whereas stable income, lower loan burden, stronger account balance, or longer credit history may reduce the predicted risk. This case-level explanation helps loan officers and decision-makers understand why a borrower is classified as high-risk or low-risk, thereby improving transparency and trust in the model.

Figure 10 illustrates how individual features contributed to a specific borrower's predicted credit risk outcome.

Furthermore, the SHAP individual plot provides a detailed explanation of how the model arrives at a prediction for a specific borrower by breaking down the

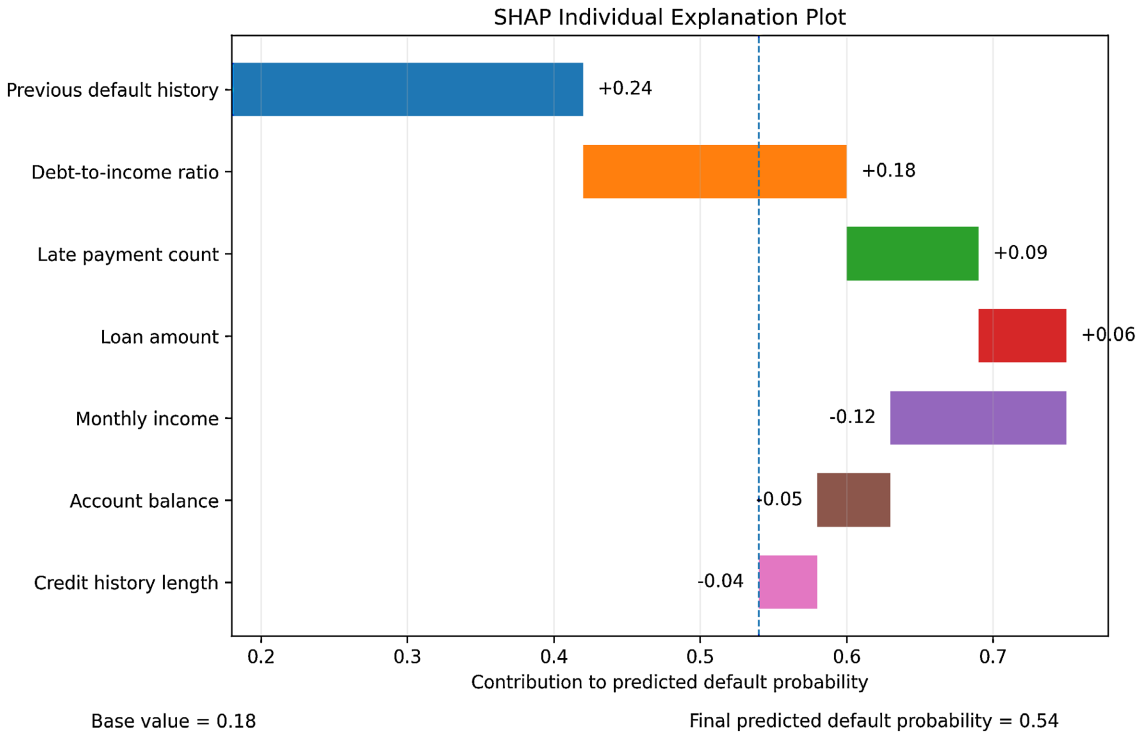


Figure 10. SHAP individual explanation plot.

contribution of each feature. Each feature either pushes the prediction toward a higher risk of default or pulls it toward a lower risk. For example, factors such as a high debt-to-income ratio, previous default history, or frequent late payments typically increase the predicted risk, as they indicate financial instability or poor repayment behavior. On the other hand, features like a stable income level or lower loan burden may reduce the risk and move the prediction toward a non-default outcome. This step-by-step contribution helps in understanding the reasoning behind the model's decision for an individual case. As a result, it becomes easier for loan officers and decision-makers to justify and trust the prediction, making the model more practical and useful in real-world banking applications.

4.8. Why the Results Improved

The improved performance of the proposed approach can be explained by several key factors that collectively enhanced the effectiveness of the model. First, the use of ensemble learning techniques, particularly XGBoost and Random Forest, played a crucial role in achieving higher accuracy. These models are well known for their ability to capture complex nonlinear relationships and interactions among features, which are often overlooked by traditional statistical methods.

Second, the quality and relevance of the selected features significantly contributed to the model's performance. The inclusion of both financial and behavioral variables, such as income level, debt-to-income ratio, and repayment history, ensured that the model was trained on realistic and meaningful indicators of credit risk. This allowed the model to better understand borrower behavior and make

more accurate predictions.

Another important factor is the application of proper data preprocessing techniques. Handling missing values, transforming categorical variables, and applying feature scaling improved the overall consistency and stability of the dataset. These steps reduced noise and ensured that the models could learn patterns more effectively.

Furthermore, hyperparameter tuning played a vital role in optimizing model performance. By carefully adjusting model parameters using techniques such as grid search, the study was able to reduce overfitting and improve generalization capability. This ensured that the model performs well not only on training data but also on unseen data. In addition, the integration of explainability techniques such as SHAP enhanced the overall reliability of the model. By providing clear insights into feature importance and prediction behavior, SHAP helped validate that the model's decisions were logical and aligned with real-world expectations. This not only improved transparency but also increased confidence in the model's outcomes.

However, it is important to note that the high performance achieved in this study may be partially influenced by the structured nature of the synthetic dataset. In real-world banking environments, data is often more complex, noisy, and imbalanced, which can reduce model performance. Therefore, the reported results should be considered as an upper-bound estimate rather than a direct reflection of real-world outcomes. Despite this limitation, the consistency of performance across multiple evaluation metrics suggests that the proposed model is robust and reliable under controlled conditions. Future research should focus on validating the model using real-world datasets to further confirm its practical applicability.

5. Business Implications

The proposed explainable machine learning framework offers substantial strategic and operational value for commercial banks, particularly in emerging financial markets such as Bangladesh. By significantly improving prediction accuracy, the framework enables financial institutions to more effectively identify high-risk borrowers, thereby reducing the likelihood of loan defaults and minimizing non-performing loans (NPLs). This directly contributes to improved financial stability and enhanced asset quality within banking portfolios. A key strength of the framework lies in its integration of explainability techniques such as SHAP and LIME, which enhance transparency in credit risk assessment. Unlike traditional black-box models, the proposed approach allows decision-makers—including loan officers, risk managers, and auditors—to understand the underlying factors influencing model predictions. This interpretability is critical in the banking sector, where regulatory compliance, auditability, and accountability are essential requirements (Noriega et al., 2023). By providing clear justifications for automated decisions, the framework supports adherence to regulatory guidelines and strengthens institutional trust. From an operational perspective, the framework

facilitates the automation of credit scoring processes, reducing reliance on manual evaluation and significantly lowering operational costs. Automated systems can process large volumes of loan applications efficiently, improving turnaround time and enhancing customer experience. Moreover, the identification of key risk drivers—such as debt-to-income ratio, income level, and repayment behavior—enables banks to refine their lending strategies, optimize credit policies, and implement more effective risk mitigation mechanisms.

In terms of financial impact, the cost-benefit analysis indicates that although the initial deployment of the framework requires investment in data infrastructure, computational resources, and technical expertise, the long-term benefits substantially outweigh these costs. Reduced default rates, improved operational efficiency, and better risk management collectively contributed to increased profitability and sustainable growth. Additionally, the framework supports data-driven decision-making, which is essential for modern banking institutions aiming to remain competitive in a rapidly evolving financial landscape. Furthermore, the adoption of explainable AI enhances organizational risk governance by reducing model risk and improving decision consistency. This is particularly important in developing economies, where banking systems are transitioning toward digital transformation and require reliable, transparent, and scalable solutions (Nallakaruppan et al., 2024).

Overall, the proposed framework not only strengthens credit risk management practices but also supports long-term financial sustainability by improving efficiency, reducing risk exposure, and enabling informed, transparent, and accountable lending decisions.

6. Strategic SWOT Analysis

To better understand the overall value and practical relevance of the proposed framework, a SWOT analysis is carried out. The study shows several important strengths. Most notably, the XGBoost model achieves high predictive accuracy, which makes the framework effective for identifying credit risk. At the same time, the use of explainable AI techniques such as SHAP and LIME improves transparency by clearly showing how decisions are made. This combination of strong performance and interpretability makes the model more reliable and suitable for real-world banking applications, where both accuracy and trust are essential. However, studying is not without limitations. One key concern is the use of a synthetic dataset, which may not fully reflect the complexity of real-world banking data. As a result, the findings may not be completely generalizable. In addition, only a limited number of machine learning models were tested, and the use of explainability methods can increase computational cost. There is also a possibility that the model's performance may decrease when applied to more complex and noisy real-world datasets.

In terms of opportunities, the proposed framework has strong potential for practical use, especially in commercial banks in Bangladesh. It can help improve

credit risk assessment, support data-driven decision-making, and enhance overall lending strategies. With further development, the model could also be integrated into automated or real-time banking systems. Future research may focus on using real-world datasets and exploring more advanced techniques, such as deep learning, to further improve performance. At the same time, some challenges and risks need to be considered. Issues such as data privacy, regulatory requirements, and possible model bias may affect the successful implementation of the system.

In addition, relying too heavily on automated decision-making could create new risks, especially in a rapidly changing financial environment. Therefore, careful monitoring and proper governance are necessary to ensure the long-term effectiveness and reliability of the framework.

7. Discussion

The findings of this study demonstrate that machine learning models, particularly XGBoost, significantly improve credit risk prediction compared to traditional methods. The results highlight the importance of using advanced algorithms to capture complex patterns in borrower behavior. From a managerial perspective, the findings provide actionable insights for banking professionals. By identifying key determinants of credit risk, such as income level, loan amount, and previous default history, managers can design more effective lending policies. Additionally, explainable predictions improve trust and transparency, enabling better communication with stakeholders.

From a business perspective, the proposed approach can support decision-making in commercial banks by identifying high-risk borrowers more accurately. This can help financial institutions reduce non-performing loans and improve lending strategies. Furthermore, the integration of explainability techniques enhances trust and transparency, which are essential for regulatory compliance and practical implementation in banking systems. This study aligns with recent research trends in explainable AI for financial applications. From an artificial intelligence perspective, this study addresses the trade-off between predictive accuracy and interpretability. By integrating explainable AI techniques with high-performance machine learning models, the proposed framework ensures both accuracy and transparency, which are essential for real-world adoption in financial systems. In the context of Bangladesh, where many banking processes are still partially manual and data availability is limited, the adoption of explainable AI can significantly enhance decision-making efficiency and support the transition toward data-driven financial systems. This is particularly important in developing countries like Bangladesh, where digital transformation in banking is still evolving, and explainable AI can enhance trust in automated decision-making systems. Since this study is based on a synthetic dataset, the findings should be interpreted as simulation-based evidence. The results demonstrate the potential usefulness of explainable machine learning for credit risk prediction in the Bangladeshi commercial banking context, but they do not confirm performance on actual borrower-

level commercial bank records. External validation using real banking data is necessary before the proposed framework can be deployed in operational credit-risk decision making. Nevertheless, the findings suggest that combining ensemble learning with explainability techniques can support transparent, data-driven, and accountable credit assessment.

8. Limitations

Despite its contributions, this study has several limitations. The main limitation is the use of synthetic data. Although the dataset was designed to reflect plausible borrower behavior and common commercial banking characteristics in Bangladesh, it may not fully capture the complexity, noise, missing-value patterns, regulatory constraints, institutional differences, and behavioral heterogeneity present in real banking data. Therefore, the reported model performance should be interpreted with caution. Second, only four machine learning models were evaluated: Logistic Regression, Random Forest, SVM, and XGBoost. Future studies may include additional algorithms such as LightGBM, CatBoost, neural networks, and hybrid ensemble models. Third, explainability techniques such as SHAP and LIME may introduce additional computational cost when applied to large-scale banking systems.

Finally, the proposed framework has not yet been validated using real borrower-level commercial bank data. Future work should focus on external validation using actual bank records, subject to ethical approval, privacy protection, and institutional permission.

9. Conclusion and Future Work

This presented an explainable machine learning framework for credit risk prediction in the context of commercial banks in Bangladesh using a synthetic borrower-level dataset. The dataset consisted of 5000 simulated loan records designed to reflect plausible demographic, financial, loan-related, and repayment-behavior characteristics associated with commercial banking practice. Logistic Regression, Random Forest, Support Vector Machine, and XGBoost were evaluated using accuracy, precision, recall, F1-score, AUC-ROC, and confusion matrix analysis. The results show that ensemble-based models outperform the baseline Logistic Regression model. XGBoost achieved the strongest overall classification performance in terms of accuracy, precision, recall, and F1-score, whereas Random Forest achieved the highest AUC-ROC value, indicating stronger class-separation ability. This distinction clarifies that XGBoost was selected as the preferred model based on overall classification performance, while Random Forest performed best in terms of discriminative capability. The integration of SHAP and LIME enhanced model interpretability by identifying the main factors influencing predicted default risk. The most influential predictors included debt-to-income ratio, monthly income, loan amount, previous default history, and late payment behavior. These explanations can support more transparent, accountable, and data-

driven credit risk assessment by helping banking professionals understand the reasoning behind model predictions. However, the findings should be interpreted with caution because the dataset is synthetic. Although the data were designed to simulate realistic borrower behavior, they may not fully capture the complexity, noise, missing-value patterns, institutional differences, and regulatory constraints of real commercial banking data. Therefore, the results should be considered simulation-based evidence rather than direct empirical evidence from actual bank records.

Future Work

Future research should validate the proposed framework using real borrower-level data from commercial banks in Bangladesh, subject to ethical approval, institutional permission, and privacy protection. Further studies may also examine cost-sensitive learning, threshold optimization, fairness analysis, additional ensemble models, fairness analysis, and real-time deployment in digital credit risk assessment systems.

Ethical Statement

The authors confirm that this study did not involve experiments on human participants or animals. No identifiable personal information, confidential borrower records, customer-level data, or private commercial bank records were used. The dataset used in this study was synthetically generated for research and simulation purposes. Therefore, the study does not involve privacy-sensitive customer information or restricted institutional data. The use of explainable artificial intelligence techniques was intended to improve transparency, interpretability, and accountability in simulated credit-risk decision-making.

Data Availability Statement

The data used in this study are synthetic and were generated for research and simulation purposes. No confidential commercial bank records, customer information, or restricted borrower-level data were used. The synthetic dataset and data-generation procedure can be made available by the corresponding author upon reasonable academic request or deposited in a public repository before final publication to support reproducibility.

Acknowledgements

The authors gratefully acknowledge the valuable guidance and scholarly support provided by Md. Saiful Islam, Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh. His constructive feedback, methodological suggestions, and careful review helped improve the clarity, structure, and overall quality of this manuscript. His advice was particularly valuable in refining the explainable machine learning framework, result interpretation, and presentation of the study. Although he is not listed as an author, the authors sincerely

appreciate his support and encouragement throughout the preparation of this research.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion, 58,* 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Bhatt, U., Andrus, M., Weller, A., & Xiang, A. (2020). Machine Learning Explain-Ability for External Stakeholders. *Proceedings of the AAAI Conference on Artificial Intelligence, 34,* 13589-13590.

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable Machine Learning in Credit Risk Management. *Computational Economics, 57,* 203-216. https://doi.org/10.1007/s10614-020-10042-0

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

Doshi-Velez, F., & Kim, B. (2017). *Towards a Rigorous Science of Interpretable Machine Learning*. arXiv:1702.08608. https://arxiv.org/abs/1702.08608

Kumar, D. (2025). Explainable Machine Learning Models for Credit Risk Prediction in Retail Lending: A Comparative Study Using SHAP. *SSRN Electronic Journal,* 13 p. https://doi.org/10.2139/ssrn.5341125

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring: An Update of Research. *European Journal of Operational Research, 247,* 124-136. https://doi.org/10.1016/j.ejor.2015.05.030

Lin, L., & Wang, Y. (2025). *SHAP Stability in Credit Risk Management: A Case Study in Credit Card Default Model*. arXiv:2508.01851

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Ed.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765-4774). Curran Associates, Inc. https://arxiv.org/abs/1705.07874

Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). Self-Published. https://christophm.github.io/interpretable-ml-book/

Nallakaruppan, M. K., Kumar, S., Kiran, P. V., & Karthikeyan, S. (2024). Credit Risk Assessment and Financial Decision Support Using Explainable Artificial Intelligence. *Risks, 12,* Article 164. https://doi.org/10.3390/risks12100164

Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine Learning for Credit Risk Prediction: A Systematic Literature Review. *Data, 8,* Article 169. https://doi.org/10.3390/data8110169

Sarder, A. A. S., Rahman, M., & Islam, M. (2024). Credit Risk Prediction Using Gradient Boosting Techniques. *Journal of Business and Management Studies, 6,* 55-66.

Seela, V. (2023). Explainable Machine Learning for Credit Risk Assessment. *International Journal of Engineering and Technology, 12,* 45-52.

Shiam, S. A. A., Hasan, M. M., & Pantho, M. J. (2024). Credit Risk Prediction Using Explainable AI. *Journal of Business and Management Studies, 6,* 61-66. https://doi.org/10.32996/jbms.2024.6.2.6

Shreya, H., & Pathak, H. (2025). *Explainable Artificial Intelligence in Credit Risk Assessment*. arXiv:2506.19383

World Bank (2019). *Global Financial Development Report: Financial Inclusion and Stability*. https://www.worldbank.org/

Ye, R., & Chen, J. (2025). *Explainable AI Framework for Credit Risk Modeling*. arXiv. https://arxiv.org/html/2506.19383v1