ASSESSING THE EFFICACY AND RISKS OF TREE-BASED ALGORITHMS IN PERSONAL DEFAULT PREDICTION: PRACTICAL INSIGHTS FROM COMMERCIAL BANKS

Nhat Minh Nguyen *, Chi Diem Ha Le **

* Ho Chi Minh University of Banking, Ho Chi Minh City, Vietnam
** *Corresponding author*, Ho Chi Minh University of Banking, Ho Chi Minh City, Vietnam
Contact details: Ho Chi Minh University of Banking, 36 Ton That Dam Street, Nguyen Thai Binh Ward, 1st District, Ho Chi Minh City 700000, Vietnam

OPEN ACCESS

How to cite this paper: Nguyen, N. M., & Le, C. D. H. (2025). Assessing the efficacy and risks of tree-based algorithms in personal default prediction: Practical insights from commercial banks. *Risk Governance and Control: Financial Markets & Institutions, 15*(3), 8–21. https://doi.org/10.22495/rgcv15i3p1

ISSN Online: 2077-4303
ISSN Print: 2077-429X

Received: 17.11.2024
Revised: 03.03.2025; 12.06.2025
Accepted: 25.06.2025

JEL Classification: C13, G21, G24
DOI: 10.22495/rgcv15i3p1

Abstract

This study evaluates the effectiveness of tree-based machine learning algorithms in predicting personal default risk, with a specific focus on commercial banks in Vietnam. By analyzing a dataset of 7,500 customers from various financial institutions collected between 2015 and 2023, we assess the performance of these algorithms using confusion matrix, accuracy, precision, sensitivity, specificity, F1 score, and area under the curve (AUC) as evaluation metrics. Our findings reveal that while traditional models like logistic regression (LR) serve as a baseline, advanced algorithms such as random forests (RF) and XGBoost (XGB) significantly enhance predictive accuracy and robustness, particularly in handling complex and imbalanced datasets (Chen & Guestrin, 2016). Among these, XGB stands out as the most effective model, demonstrating superior performance across all evaluation metrics (Li et al., 2020). Additionally, the feature importance analysis highlights the critical roles of loan characteristics, applicant financial information, employment and residential information, and credit history in default prediction. Notably, loan term, highest credit cap, employment tenure, and active accounts number emerge as the most influential features, shaping the individual probability of default. However, limitations in data availability and the directional impact of feature variables within the model may reduce the generalizability and interpretability of the predictive model. This research provides valuable insights for financial institutions aiming to improve their credit risk management practices by adopting sophisticated machine-learning models to predict personal defaults.

Keywords: Personal Default Prediction, Tree-Based Machine Learning Algorithms, XGBoost, Credit Risk Management, Commercial Banks

Authors' individual contribution: Conceptualization — N.M.N. and C.D.H.L.; Software — N.M.N.; Validation — N.M.N. and C.D.H.L.; Formal Analysis — C.D.H.L.; Investigation — N.M.N. and C.D.H.L.; Resources — N.M.N. and C.D.H.L.; Data Curation — N.M.N. and C.D.H.L.; Writing — Original Draft — N.M.N. and C.D.H.L.; Writing — Review & Editing — C.D.H.L.; Visualization — N.M.N. and C.D.H.L.; Supervision — N.M.N.; Project Administration — N.M.N.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest

Acknowledgement: The Authors thank all lecturers and managers of Ho Chi Minh University of Banking (HUB). This article is supported and funded by HUB, Vietnam.

1. INTRODUCTION

Accurate estimation of personal default risk is a cornerstone of modern credit risk management, enabling financial institutions to comprehensively assess borrowers' financial conditions, determine appropriate interest rates, and establish well-informed lending terms. Beyond individual risk assessment, accurate predictions also strengthen loan portfolio management by enhancing credit quality evaluation, optimizing pricing strategies, and mitigating financial instability. This capability ultimately contributes to the development of more resilient credit risk management frameworks, helping financial institutions minimize potential losses and improve operational efficiency (Eyalsalman et al., 2024; Arifaj & Baruti, 2023).

Several machine learning algorithms, including logistic regression (LR), decision trees (DT), random forests (RF), and XGBoost (XGB), have been found to greatly enhance default prediction accuracy. These advanced strategies outperform existing methods for dealing with complex data and capturing nonlinear relationships between variables (Bao et al., 2019). LR is a well-known statistical procedure that is both simple and easy to understand. It works especially well when there is a linear relationship between the log-odds of the dependent variable and the independent variables (Peng et al., 2002). On the other hand, DT is an intuitive, tree-structured model that divides data into segments depending on feature values, making them simple to visualize and comprehend (Lee et al., 2022). RF, an ensemble approach, increase decision tree performance by averaging many trees' outputs, reducing overfitting and increasing prediction accuracy (Breiman, 2001). Meanwhile, XGB, a gradient boosting algorithm, has grown in prominence due to its ability to handle nonlinear connections robustly and provide improved optimization capabilities (He et al., 2024).

However, numerous academics have expressed differing opinions on the efficacy of these models. Pate et al. (2023) argue for LR since it is simple and successful in linear circumstances. However, Dumitrescu et al. (2022) argues that LR's performance deteriorates with non-linear data and complicated interactions, implying that even with additional interaction terms and polynomial characteristics, the model's interpretability is severely reduced. In contrast, Chen and Guestrin (2016) argue that XGB is preferable at dealing with big, complex, and unbalanced datasets. Li and Chen (2020) back this up by showing that XGB routinely outperforms other models, including LR and RF, in terms of accuracy and area under the curve (AUC). They suggest that RF strikes a fair balance between simplicity and predictive capability, making it a dependable option for complex datasets. Deng and Runger (2013) discovered that RF outperforms both LR and DT in datasets with high dimensionality and complexity. Recent investigations back up these conclusions. According to Velarde et al. (2024), XGB surpasses other models in predicting accuracy, especially when dealing with complicated relationships and unbalanced datasets. Furthermore, Shetty et al. (2022) found that RF outperformed DT and LR in forecasting loan defaults, demonstrating the ensemble method's resilience and generalization capabilities.

Moreover, debate about which feature group most significantly impacts personal default risk continues, with various perspectives highlighting the complexity of credit risk assessment (Goel & Rastogi, 2023). Some researchers argue that loan characteristics like loan duration and purpose are crucial, as they directly affect financial commitments and may increase default risk over longer periods (Qi, 2023). Conversely, other studies suggest that employment and residential stability are better predictors of default risk, reflecting long-term financial health and stability (Naili & Lahrichi, 2022). Research also points to the importance of applicant financial information, such as credit limits and income levels, as critical indicators of creditworthiness. However, an over-reliance on any single factor, such as credit history, might fail to capture the full picture of a borrower's financial situation, suggesting the need for a balanced approach that integrates various aspects of a borrower's profile in risk assessment (Avery et al., 2004).

Therefore, the purpose of this study is to provide an in-depth comparison of tree-based algorithms such as DT, RF, and XGB in predicting personal default risk. By identifying the most effective forecasting model, the article aims to determine the most important features influencing personal default risk. The results are expected to help financial institutions gain deeper insights into the performance of these algorithms, helping them make more accurate decisions in managing credit risk.

The rest of the article is structured as follows: Section 2 provides the theoretical background, reviewing the LR, DT, RF, and XGB algorithms along with their application in credit risk assessment. Section 3 describes the research methodology, including research design, tree-based algorithms, and performance metrics. Section 4 presents experimental results, comparing the performance of the models based on different evaluation criteria. Finally, Section 5 summarizes the key findings, discusses the benefits and limitations of tree-based models in personal default prediction, and suggests future research directions.

2. LITERATURE REVIEW

Advanced models like neural networks, evolutionary algorithms, and ensemble methods are increasingly adopted by banks and financial institutions to enhance the accuracy of credit assessments and decision-making processes (Addo et al., 2018). LR remains a popular method for binary classification tasks, such as predicting loan defaults. Liu et al. (2024) demonstrate LR's effectiveness with linear relationships. However, its limitations become evident in non-linear and complex data structures. Fitzpatrick and Mues (2021) highlight that while interaction terms can enhance LR's predictive capabilities, they may increase model complexity without significant accuracy improvements in complex scenarios. Pan et al. (2024) proposed that the efficiency of LR can be improved by reducing the dimensionality of the data using the principal component analysis technique. DT is valued for its simplicity and transparency, making it useful for initial data exploration and feature selection (Lee et al., 2022). However, Sagi and Rokach (2018) warn against their propensity to overfit, especially with noisy data. Ensemble methods like RF can mitigate these risks and improve model resilience (Sun et al., 2024).

RF addresses the overfitting issue inherent in DT by averaging the results of multiple trees, thereby increasing robustness and generalizability. Breiman (2001) and Iranzad and Liu (2024) demonstrated that RF reduces variance and effectively handle high-dimensional data. Ghatasheh (2014) found that RF outperforms both LR and DT in complex datasets, making it a robust choice for various credit risk prediction tasks.

XGB is renowned for its exceptional performance with large, complex, and imbalanced datasets. Chen and Guestrin (2016) emphasize XGB's ability to handle missing data and capture complex variable interactions. Studies by Uddin and Rahman (2017) and Suhadolnik et al. (2023) consistently show XGB outperforming models like LR and RF in terms of accuracy and AUC, establishing it as a preferred choice for credit risk management.

Recent research supports these findings. Zhou (2023) notes that RF outperforms DT and LR in forecasting loan defaults, highlighting ensemble methods' resilience and generalizability. Barua et al. (2021) compared XGB with CatBoost, finding XGB superior in several instances despite CatBoost's promising results. However, in the study of Nguyen and Ngo (2025), it was shown that XGB did not show superiority over LightGBM and CatBoost in predicting customer bankruptcy risk.

While LR is praised for its simplicity and interpretability, it often falls short in complex and non-linear situations compared to advanced models like XGB and RF. XGB stands out as the most robust and accurate model for predicting personal loan defaults, particularly in challenging datasets, though RF also performs well due to its resilience and capability to handle complex data structures effectively. The ongoing debate about which features most significantly impact personal default risk highlights the complexity of credit risk assessment. Researchers are divided on the importance of different variables in predicting defaults.

Some studies emphasize the critical role of loan characteristics, such as duration and purpose, as they directly affect financial commitments and lenders' risk exposure. Prolonged loan terms may increase default risk due to extended financial obligations (Qi, 2023). Alternatively, other studies argue that employment stability and socio-economic factors are more accurate predictors of default risk, reflecting long-term financial health and stability. Stable employment and homeownership are associated with lower default risks, indicating financial stability (Naili & Lahrichi, 2022). Applicant financial information, including credit limit, income, and debt levels, is also crucial in evaluating creditworthiness. Higher income generally correlates with lower default risk, though focusing solely on financial metrics may overlook broader socio-economic factors (Jiang, 2022). Credit history provides insights into financial behavior and reliability, but over-reliance on it can ignore current financial situations (Avery et al., 2004).

Other critical factors include job stability and asset ownership are consistently linked to better financial health and lower default rates (Fay et al., 2002). Additionally, the loan's purpose, monthly obligations, credit history duration, and number of open accounts are important. Income-generating loans and well-managed credit profiles reduce default risk (Qi, 2023). Longer intervals since the last default and smaller credit balances indicate better credit behavior (Hussin Adam Khatir & Bee, 2022). Effective management is crucial to mitigating these risks, despite the financial benefits of large credit limits (Lu et al., 2024).

The review of the research results above indicates that there are still certain debates about the effectiveness of tree-based algorithms. Additionally, determining which feature groups play an important role in predicting personal default risk remains an issue that requires further clarification.

3. METHODOLOGY

3.1. Research design

The study focusses on creating and testing multiple machine learning models for forecasting personal loan defaults, using a systematic approach from data preparation to model deployment. The models being considered include XGB, RF, DT, and LR. The methodological framework follows a structured process to ensure data integrity, model robustness, and reliable evaluation metrics. The research design is illustrated in Figure 1, outlining key analytical stages.

The first stage is data preparation, ensuring the dataset is of high quality and suitable for training predictive models. The dataset, which comprises financial and credit information from Vietnam's commercial banks, undergoes multiple preprocessing steps:

1. *Data cleaning*: Duplicate entries are identified and removed to prevent redundancy and ensure analysis integrity. Missing values are handled using multiple imputation techniques, such as mean, mode, or median imputation, depending on the distribution of missing data. For non-normally distributed variables, K-nearest neighbors imputation is applied to preserve data structure.

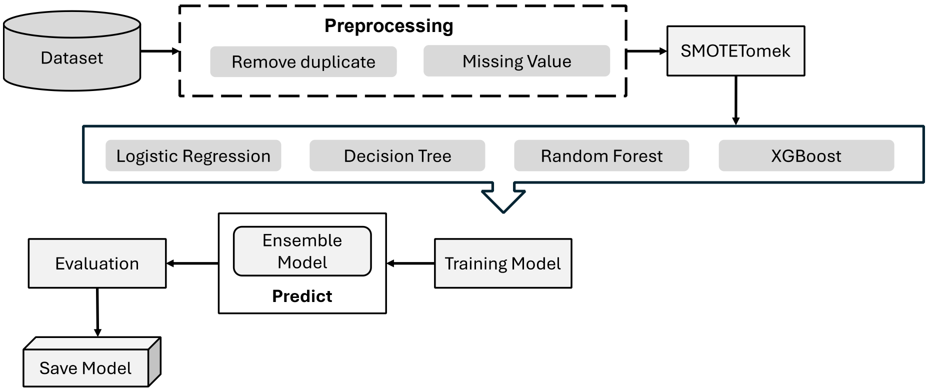
2. *Handling class imbalance*: Since loan default datasets often exhibit imbalance, the synthetic minority over-sampling technique (SMOTE) with Tomek links is used. SMOTE generates synthetic examples for the minority class, while Tomek links remove ambiguous borderline samples from the majority class, leading to a more distinct decision boundary and reducing overfitting to the dominant class.

The next stage is model training phase involves developing and fine-tuning the four selected models. Each model is initialized with baseline hyperparameters, which are optimized using an iterative tuning process. During training, the models iteratively alter their parameters to minimise the loss function, and the process is continuously monitored to guarantee effective learning while avoiding overfitting. A validation set is used to fine-tune each model's hyperparameters. Grid search and random search are used to find the ideal collection of hyperparameters that maximise the model's performance on the validation set.

After training, model evaluation is conducted on a separate test set using seven performance metrics such as confusion matrix, accuracy, sensitivity, specificity, precision, F1 score, and AUC. These measures give a thorough evaluation of each model's ability to accurately forecast personal loan defaults.

Finally, the deployment phase ensures that trained models are preserved and ready for real-world application. The models are serialised and saved in a way that allows for easy reloading and use in future forecasts. This includes preserving the model's architecture, weights, and any necessary preprocessing processes. This project intends to create strong prediction models that can help financial institutions identify persons at risk of default by utilising advanced data preparation techniques and cutting-edge machine learning algorithms. This comprehensive approach guarantees that the models remain accurate, reliable, and ready for practical deployment.

Figure 1. Description of research design



Source: Gao et al. (2021).

3.2. Forecasting methods based on tree-based machine learning models

The class of machine learning models based on tree structures includes methods that use hierarchical tree topologies to perform supervised learning tasks such as classification and regression. These models work by splitting the dataset into increasingly smaller subgroups using decision rules generated from attribute feature values. Tree-based algorithms are regarded as simple, interpretable, and capable of handling a wide range of data types. Popular models in this field include DT, RF, XGB. The extensive use of these models can be ascribed to their ability to uncover important patterns and make accurate predictions from complicated datasets.

3.2.1. Decision tree

DT is an essential classification tool in machine learning, particularly useful for identifying personal default risk based on customer attributes. In predicting personal bankruptcy, DT utilizes metrics such as entropy or the Gini index to optimize classification, effectively distinguishing between high-risk and low-risk customers. This research selects the Gini index due to its ease of calculation and its ability to make DT more computationally efficient. This approach enhances the precision and effectiveness of identifying individuals likely to default, providing valuable insights for financial institutions managing credit risk.

The Gini index, also referred to as Gini impurity, is a metric used to quantify the likelihood of a randomly chosen sample being misclassified if it were assigned a label at random based on the label distribution within that subset. The Gini index is defined for a dataset S as follows:

$$G(S) = 1 - \sum_{i=1}^{n} p_i^2 \tag{1}$$

In this measure, p_i represents the proportion of samples that belong to class i within set S (in this case, class i could be "default" or "non-default"). The lower the Gini value, the more homogeneous the dataset becomes.

In DT, the selection of features for data splitting at each node is guided by the objective of achieving the lowest Gini impurity post-split. The *Gini gain*, which quantifies the reduction in Gini impurity, is a crucial metric used in the construction of DT to evaluate the improvement (or reduction) in impurity following the division of data based on a specific feature. The *Gini gain* is computed by subtracting the weighted average of the Gini impurities of the subsets resulting from the split from the initial Gini impurity of the dataset:

$$Gini\ gain\ (S, A) = G(S) - \sum_{v\ \epsilon\ Value(A)} \frac{|S_v|}{|S|} G(S_v) \tag{2}$$

In this context, S_v is the subset of S when the feature A takes the value v.

Gini gain measures the extent to which each feature contributes to reducing heterogeneity within the dataset when employed to split data at a node. The feature with the highest *Gini gain* is deemed the most optimal for forming the subsequent split node, as it generates subsets with the highest degree of homogeneity.

The tree-building process is iteratively repeated across the data features until one of the following conditions is met: 1) each leaf node attains a specified level of homogeneity, or all transactions at a node belong to a single class, either "default" or "non-default" (Gini = 0); 2) the tree reaches a predefined depth; 3) the number of transactions at a node falls below a certain threshold. This systematic methodology ensures that the decision tree optimally categorizes the data, minimizing impurity at each stage of the split.

3.2.2. Random forest

RF is an ensemble technique that combines multiple DT to reduce the risk of overfitting, enhance accuracy, and improve the model's generalizability. In the context of predicting personal bankruptcy, a RF consists of several DT $T_b(X)$, where each tree b is a function of the training dataset X, constructed from a bootstrap sample of X. This approach leverages the power of multiple trees to produce a robust predictive model that is less prone to the biases and variances inherent in individual DT. The number of trees B and the operational methodology of each tree within the forest are defined as follows:

$$RF = \{T_1(X), T_2(X), \dots, T_B(X)\} \quad (3)$$

In the RF model, each tree $T_b(X)$ is constructed following these guidelines:

1) each tree T_b is developed from a bootstrap sample of the original dataset X, denoted as X_b (where X_b = *BootstrapSample*(X));

2) during the construction of each node within the tree, a random subset m of the attributes is selected from the total p attributes of the dataset ($m \leq p$). These attributes are then evaluated based on the Gini impurity index to determine the optimal splitting point.

In the context of predicting personal bankruptcy, the prediction made by the RF model for a new customer x is determined by aggregating the majority vote from the ensemble of trees. This collective decision-making process leverages the strengths of multiple DT to enhance the accuracy and reliability of predicting default risk. By combining the predictions from several trees, the RF model mitigates the influence of individual tree biases and variances, resulting in a more robust and dependable assessment of an individual's likelihood to default.

$$\hat{y} = model\{T_1(x), T_2(x), \dots, T_B(x)\} \quad (4)$$

3.2.3. XGBoost

XGB is an advanced variant of gradient boosting machines (GBM) designed to optimize performance and speed while effectively managing large-scale data, particularly in predicting personal bankruptcy. XGB incorporates several technical enhancements to improve efficiency and practical applicability in this context.

XGB extends the capabilities of GBM by introducing a more precise loss function and more efficient optimization techniques. It typically employs a log-likelihood loss function, similar to GBM, but includes additional regularization components to enhance model training and prevent overfitting. These enhancements contribute to its robustness and accuracy in predicting the likelihood of personal default, making it a powerful tool for financial institutions looking to assess credit risk and manage potential defaults effectively.

$$L(y, f(x)) = \sum_{i=1}^{n} l(y_i, f(x_i)) + \sum_{k=1}^{K} \Omega(f_k) \quad (5)$$

where,

- $l(y_i, f(x_i))$ represents the logistic loss function;
- $\Omega(f_k)$ denotes the regularization function for each tree f_k in the model, typically including both L1 (lasso) and L2 (ridge) regularization:

$$\Omega(f) = \gamma T + ½\lambda \sum_{j=1}^{T} w_j^2 \quad (6)$$

with γ and λ are the regularization parameters and T is the number of leaf nodes in the tree, w_j is the value at the leaf node.

XGB offers enhanced control through its regularization parameters, allowing its models to better resist overfitting and typically deliver superior performance on large and complex datasets. This capability ensures that XGB remains highly effective in diverse analytical scenarios, particularly when managing data-rich environments related to personal bankruptcy prediction. Such control is instrumental in maintaining model robustness and accuracy, making XGB a preferred choice for advanced predictive modeling tasks involving credit risk assessment and default prediction. Its ability to handle large datasets with complex patterns makes it a valuable tool for financial institutions seeking to accurately identify individuals at risk of default.

3.2.4. Logistic regression

LR is not a tree-based machine learning model, it is a linear classification model that uses sigmoid function to estimate the probabilities of classes. With the advantages of simplicity, ease of interpretation, and great suitability for binary classification problems such as default risk prediction, the LR model has become the standard for efficiency that other models need to surpass.

Assuming $X_1,\dots X_n$ are a set of features reflecting the financial characteristics of an individual, the probability of personal default ($Y = 1$) will be determined through the sigmoid function as follows:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (7)$$

where,

- $P(Y = 1|X)$ is the probability of personal default, which ranges from 0 to 1;
- β_0 is the intercept coefficient;
- $\beta_1, \beta_2, \dots \beta_n$ are the regression coefficients that reflect the influence of the feature variables $X_1,\dots X_n$ on the probability of personal default.

3.3. Performance measures for evaluating the model's ability to predict default risk

Models for predicting defaults are evaluated using a variety of performance measures, including the confusion matrix, accuracy, sensitivity, specificity, precision, F1 score, and AUC:

3.3.1. Confusion matrix

The confusion matrix presented in Table 1 is an important tool for assessing the performance of classification models. It provides a complete examination of the model's accurate and wrong predictions, allowing for a better understanding of its performance across several classes. The confusion matrix, which compares projected values to actual outcomes, indicates the model's strengths and places for development, which is especially useful in financial applications like estimating default risk. This thorough analysis assists in identifying particular areas where the model excels or requires development, allowing for better informed credit risk management decisions.

Table 1. Confussion matrix

		Predicted class	
	Classes	***Non-default***	***Default***
Actual class	Non-default	*True negative* (*TN*) Instances where the model predicts that the enterprise will not default, and it indeed does not default	*False positive* (*FP*) Instances where the model predicts default, but in reality, the enterprise does not default.
	Default	*False negative* (*FN*) Instances where the model predicts that the enterprise will not default, but in reality, it does default.	*True positive* (*TP*) Instances where the model predicts that the enterprise will default and, in fact, it does default.

Source: Ying (2018).

3.3.2. Other measurement indicators: accuracy, sensititvity, specificity, precision, F1 score, AUC

Other important criteria measuring the effectiveness of models, such as accuracy, sensitivity, specificity, precision, F1 score, and AUC, are clearly described in Table 2 as follows:

Table 2. Groups of indicators to evaluate the reliability of a model

Indicators	***Definition***	***Formula***
Accuracy ratio	The overall accuracy of the model is determined by comparing all accurate predictions to the total predictions.	$Accuracy = \frac{TN + TP}{TN + FN + TP + FP}$
Sensititvity ratio	This indicates that the model can accurately identify default consumers.	$Sensitivity = \frac{TP}{TP + FN}$
Specificity ratio	Evaluates the model's ability to appropriately detect non-default clients.	$Specificity = \frac{TN}{TN + FP}$
Precision ratio	Out of all projected default customers, this metric reflects the model's accuracy in predicting customers.	$Precision = \frac{TP}{TP + FP}$
F1 score	Balances precision and sensitivity, resulting in a single score that takes into account both erroneous positives and false negatives.	$F1\ score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$
AUC	AUC, or area under the curve, measures the area beneath the receiver operating characteristic (ROC) curve and depicts the trade-off between true positive and false positive rates.	

Source: Ying (2018).

4. RESEARCH RESULTS

4.1. Data research

The dataset utilized in this study to predict personal defaults comprises data from 7,500 customers sourced from several commercial banks and financial institutions across Vietnam, collected over the period from 2015 to 2023. It is provided by the Finance and Banking Research and Training Center at Ho Chi Minh University of Banking, which supports both research and educational initiatives. The dataset includes encrypted customer data to adhere to data protection laws, encompassing financial details and credit statuses at the time of data gathering.

Table 3. Dataset features

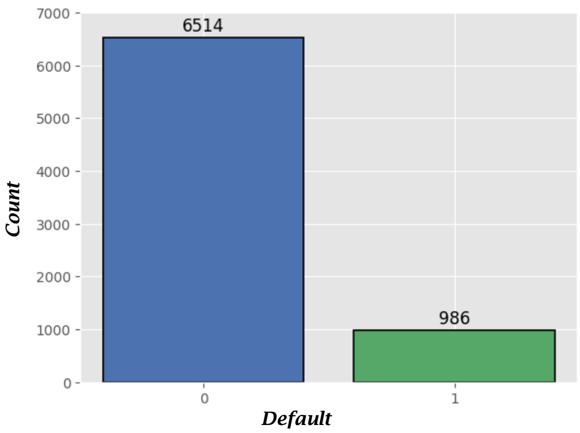
Variables	***Groups***	***Variables***	***Description***
Explanatory variables	Loan characteristics	Principal amount (*LC1*)	The amount currently approved for the loan
		Loan term (*LC2*)	The term length of the loan (short-term or long-term)
		Financing objective (*LC3*)	The reason for taking the loan
	Applicant financial information	Annual earnings (*FI1*)	The annual income of the applicant
		Monthly obligations (*FI2*)	The total monthly debt obligations
		Highest credit cap (*FI3*)	The highest amount of credit ever extended to the applicant
		Outstanding credit (*FI4*)	The current outstanding balance on all credit accounts
	Employment and residential information	Employment tenure (*ER1*)	The number of years the applicant has held their current job
		Residential status (*ER2*)	The type of home ownership (own, mortgage, or rent)
	Credit history	Credit tenure (*CH1*)	The total number of years the applicant has had credit
		Months since last default (*CH2*)	The number of months since the applicant last missed a payment
		Active accounts number (*CH3*)	The total number of open credit accounts
Target variable		Default	Indicates whether the applicant is currently in default or not

Source: Authors' elaboration.

The dataset consists of 12 feature variables divided into four groups: loan characteristics, applicant financial information, employment and residential information, and credit history, as detailed in Table 3. Besides, the target variable, *Default*, described in Figure 2, is divided into two categories: 0 (non-default) and 1 (default).

The default rate is 13.15%, which is typical of default risk prediction datasets because there are fewer defaulting clients than non-defaulting ones. This is clearly illustrated in the provided bar chart and summary table provided in Figure 2, where 86.85% of the dataset (6,514 instances) are non-defaults, and only 13.15% (986 instances) are defaults. Such a pronounced class imbalance underscores the challenge of accurately predicting defaults and necessitates specialized techniques to address potential biases in predictive models.

Figure 2. Default distribution



Source: Authors' calculations.

Table 4. Client default and non-default rates in the dataset

Default	*Count*	*Percentage*
0.0	6514	86.65%
1.0	986	13.15%

The dataset was thoroughly analysed with Python and powerful machine learning frameworks. XGB, known for its speed and accuracy, was created using 'XGBClassifier()', which handles complicated data with ease. The 'DecisionTreeClassifier()' and 'RandomForestClassifier()' functions were used, both of which are capable of managing big datasets and categorical data. LR, implemented using 'LogisticRegression()', was used as a basic comparison, emphasizing its simplicity and interpretability. These models jointly produced reliable forecasts for personal loan defaults, providing vital insights for credit risk management. Moreover, to guarantee robustness and reliability, the dataset was divided into training (90%) and testing (10%) subsets, allowing for thorough training and effective evaluation of the models' generalization capabilities. These strategies, when paired with thorough data preparation and robust model training and assessment, seek to improve forecast accuracy and give significant insights for financial institutions controlling credit risk.

4.2. Comparison results on the predictive ability of the models

The out-of-sample test results of the boosting algorithms including XGB, RF, DT, and LR are presented in detail in Table 5.

Table 5. Default prediction results of models on out-of-sample data sets

No.	*Model*	*Accuracy*	*Precision*	*Sensitivity*	*Specificity*	*F1 score*	*AUC*
0	Logistic regression (LR)	0.699920	0.700754	0.699920	0.668262	0.699622	0.771430
1	Decision tree (DT)	0.855547	0.855658	0.855547	0.846890	0.855537	0.855554
2	Random forest (RF)	0.925778	0.925788	0.925778	0.928230	0.925778	0.981775
3	XGBoost (XGB)	0.947326	0.947399	0.947326	0.953748	0.947324	0.988525

Source: Authors' calculations.

Accuracy assesses the model's overall accuracy. LR has the lowest accuracy (69.99%), implying that it accurately predicts defaults and non-defaults less frequently than other models. DT improves greatly, with an accuracy of 85.55%, indicating a large gain in total correctness. RF improves performance with an accuracy of 92.58%, indicating its resilience. XGB has the best accuracy at 94.73%, making it the most trustworthy model for correct predictions.

Precision is defined as the ratio of accurately predicted positive observations to all expected positives. LR has a precision of 70.08%, indicating that it is moderately good at accurately predicting defaults among all projected defaults. DT outperforms this with an accuracy of 85.57%, lowering the number of false positives. RF outperforms with a precision of 92.58%, suggesting great reliability in labeling defaults. XGB has the best precision (94.74%), demonstrating its excellent ability to eliminate false positives efficiently.

Sensitivity assesses the model's ability to recognize all instances of the default class. LR has a sensitivity of 69.99%, showing that it fails to find default situations. DT improves significantly with a sensitivity of 85.55%, increasing its capacity to detect defaults. RF performs remarkably well, with a sensitivity of 92.58%, showing great efficacy in detecting genuine default scenarios. XGB has the maximum sensitivity (94.73%), making it the most effective model for reducing financial losses caused by unreported defaults.

Specificity refers to the fraction of genuine negatives that are accurately detected. LR has a specificity of 66.83%, showing modest efficacy in detecting non-default instances. DT improves with a specificity of 84.69%, indicating higher effectiveness in distinguishing non-defaults. RF also performs well with a specificity of 92.82%, demonstrating a good capacity to eliminate false positives. XGB excels with the highest specificity (95.37%), indicating higher efficacy in reliably recognizing non-default scenarios.

The F1 score is the harmonic mean of precision and sensitivity, offering a single metric that balances both objectives. LR has an F1 score of 69.96%, indicating modest competence in both accuracy and sensitivity. DT increases greatly with an F1 score of 85.55%, showing a better mix of precision and sensitivity. Random forest earns an exceptional F1 score of 92.58%, indicating that it is a well-balanced and successful model overall. XGB has the greatest F1 score (94.73%), confirming its status as the most balanced and successful model in this investigation.

AUC indicates the model's ability to differentiate across classes. It is very useful for testing binary classification models. LR has an AUC of 77.14%, showing that it is moderately good at discriminating between default and non-default scenarios. DT improves with an AUC of 85.55%, suggesting higher discrimination abilities. With an AUC of 98.18%, RF outperforms other classification methods. However, XGB has the greatest AUC (98.85%), demonstrating its greater ability to differentiate between default and non-default scenarios.

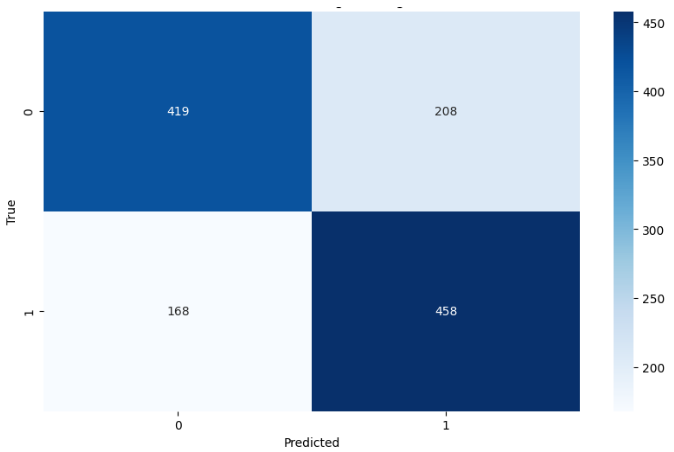
These measures demonstrate clear disparities in model performance. LR, while straightforward and easy to understand, is the least successful model, particularly in terms of accuracy and sensitivity. DT improves significantly, achieving a better balance between detecting defaults and non-defaults. However, the advanced models, RF and XGB, surpass LR and DT on all measures. RF strikes a good balance, with excellent accuracy, precision, sensitivity, specificity, and AUC, making it a solid choice for predicting personal defaults. However, XGB outperforms all other models, with the greatest metrics across the board, making it the most successful model for default prediction in this research.

Several research papers back up the conclusions of this analysis. For example, Chen and Guestrin (2016) emphasized XGB's efficiency and good performance in classification tasks, displaying a higher AUC than other models. Similarly, Gumus and Kiran (2017) discovered that XGB frequently outperforms other machine learning algorithms for predicting defaults, especially in financial scenarios. Chen and Guestrin (2016) found that XGB's scalability and flexibility enable it to easily handle huge datasets, resulting in greater accuracy and AUC scores. They demonstrated that XGB's ability to optimize trees using a gradient boosting framework produces superior performance metrics than typical machine learning models. This is especially significant in financial applications because differentiating between default and non-default scenarios requires great precision. Barboza et al. (2017) conducted a thorough analysis of several machine-learning models for default prediction. Their findings revealed that XGB not only had the greatest accuracy and AUC but also displayed resilience across several datasets. The study emphasized that XGB's regularization approaches serve to reduce overfitting, making it a dependable option for financial risk modeling.

Moscatelli et al. (2020) provide more supportive evidence by comparing the effectiveness of several machine-learning models in forecasting business defaults. Their findings revealed that XGB regularly outperformed other models in terms of both precision and sensitivity. They ascribed XGB's higher performance to its capacity to detect complicated patterns in data using boosting rounds. Additionally, Li et al. (2020) investigated the use of XGB in credit rating and default prediction. Their findings revealed that XGB's sophisticated boosting algorithms improve predicted accuracy and the management of unbalanced datasets. They remarked that XGB's feature significance scores give useful insights into the elements impacting default risk, allowing for better-informed decision-making.

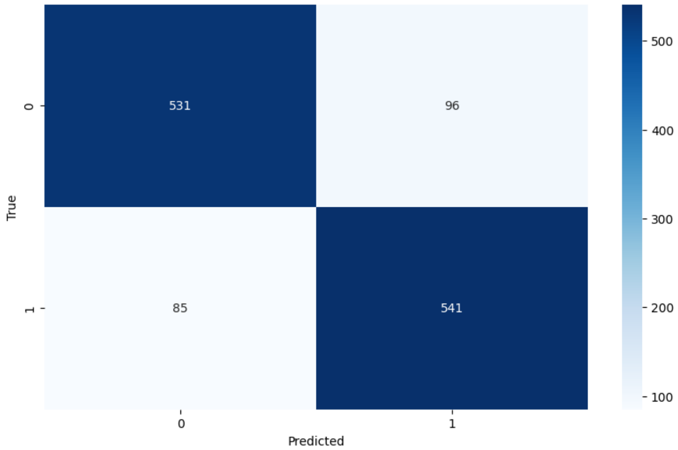
The study of the confusion matrices in Figure 3 clearly indicates the XGB model's improved performance in forecasting personal loan defaults when compared to LR, DT, and RF. The LR model produced 419 accurate predictions for non-defaults and 458 correct predictions for defaults, as opposed to 208 erroneous predictions for non-defaults and 168 incorrect predictions for defaults. The DT model improved, with 531 accurate predictions for non-defaults and 541 for defaults, compared to 96 wrong predictions for non-defaults and 85 for defaults. RF did even better, with 582 right predictions for non-defaults and 578 correct predictions for defaults, as well as fewer wrong predictions (45 for non-defaults and 48 for defaults). However, the XGB model outscored all other models, with 598 correct non-default predictions and 589 for defaults. It had the fewest erroneous predictions, with just 29 for non-defaults and 37 for defaults. In conclusion, XGB stands out for its greater capacity to handle complicated datasets and deliver the most accurate predictions with the fewest mistakes.

Figure 3a. Prediction results of models on the confusion matrix: Logistic regression



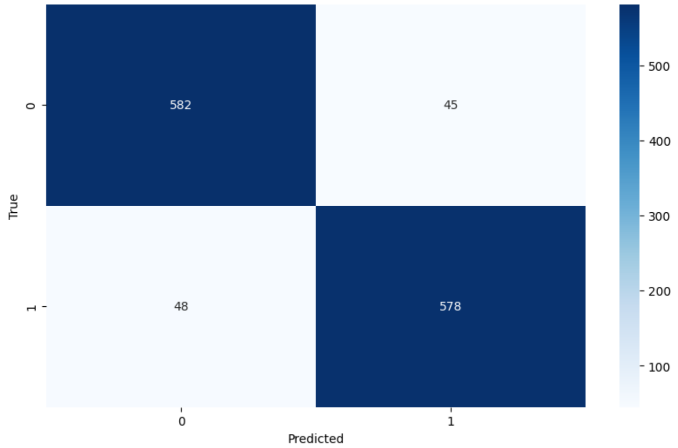
Source: Authors' calculations.

Figure 3b. Prediction results of models on the confusion matrix: Decision tree



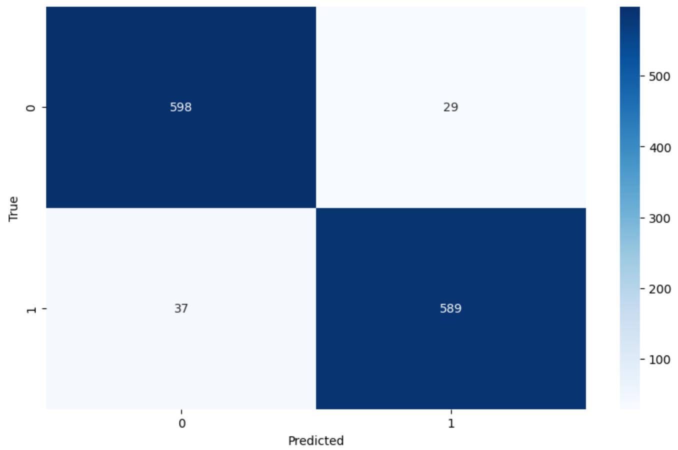
Source: Authors' calculations.

Figure 3c. Prediction results of models on the confusion matrix: Random forest



Source: Authors' calculations.

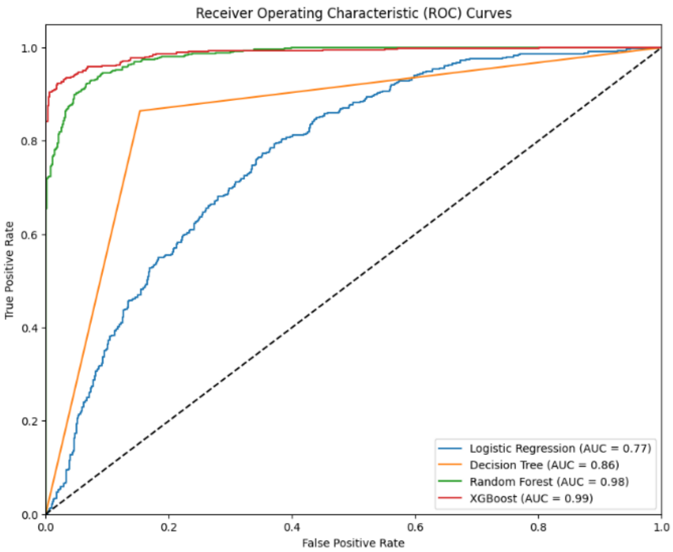
Figure 3d. Prediction results of models on the confusion matrix: XGBoost



Source: Authors' calculations.

Figure 4 shows the ROC curve and AUC data, which clearly reflect the XGB model's improved performance. XGB scored an AUC of 0.99, showing that it has remarkable classification capabilities and can reliably discriminate between default and non-default instances. In comparison, the RF model performed wonderfully, with an AUC of 0.98, closely matching XGB in terms of performance. The DT model performed well, with an AUC of 0.86, but fell behind the ensemble techniques. LR showed the lowest performance of the models, with an AUC of 0.77, indicating its limits in dealing with complicated, non-linear connections in the data.

Figure 4. Prediction results of models on the ROC curves



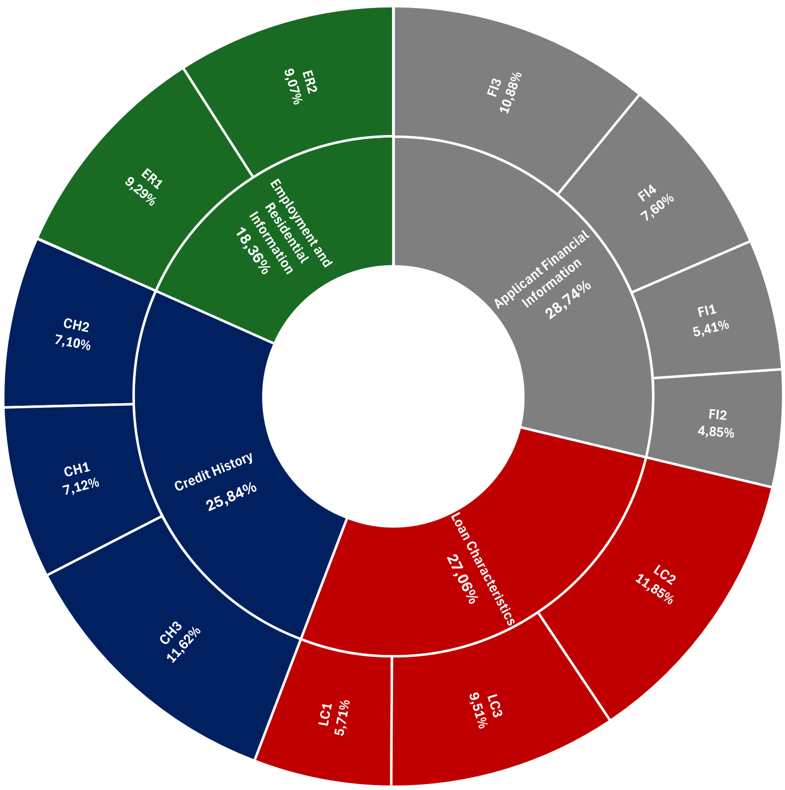
Source: Authors' calculations.

After comparing the performance of models in predicting personal default, the XGB model emerged as the most outstanding, completely outperforming the others across all evaluation criteria. Therefore, to gain a more profound understanding of how the XGB model operates, the next section will focus on discussing the importance of the selected features in this model.

4.3. Discussion on feature importance in the XGB model

Zheng et al. (2017) demonstrated that understanding feature importance values provides valuable insights into how each feature impacts the prediction results, which helps in comprehending the decision-making process of the model. In Figure 5, the analysis highlights the significant importance of each feature group in predicting personal default risk through the XGB model.

Figure 5. Feature importance values in predicting personal default risk



Source: Authors' calculations based on the study of Zheng et al. (2017).

Table 6. Ranking of feature importance values from high to low

Variables	*Feature importance values*
Loan term (*LC2*)	11.85%
Active accounts number (*CH3*)	11.62%
Highest credit cap (*FI3*)	10.88%
Financing objective (*LC3*)	9.51%
Employment tenure (*ER1*)	9.29%
Residential status (*ER2*)	9.07%
Outstanding credit (*FI4*)	7.60%
Credit tenure (*CH1*)	7.12%
Months since last default (*CH2*)	7.10%
Principal amount (*LC1*)	5.71%
Annual earnings (*FI1*)	5.41%
Monthly obligations (*FI2*)	4.85%

First, the applicant financial information group holds the largest weight at 28.74%, indicating that financial capabilities such as annual earnings, monthly obligations, highest credit cap, and outstanding credit are critical factors in assessing default risk. Second, the loan characteristics group accounts for 27.06%, emphasizing that factors such as loan term, financing objective, and principal amount all impact the borrower's ability to repay. The credit history group, with 25.84%, also plays a vital role, as elements like the number of months since the last default, the number of active accounts, and credit tenure are closely linked to predicting credit risk. Finally, the employment and residential information group, at 18.36%, shows that employment tenure and residential status are also influential factors affecting debt repayment capability. This analysis demonstrates that grouping features by similar characteristics not only organizes data systematically but also provides insights into how these features interact and affect personal default risk.

The feature importance values provide insights into how different characteristics influence the prediction of personal default risk. These values are ranked based on their contribution to the predictive model, offering a clear picture of which features are most impactful.

Loan term (11.85%): This feature, from the loan characteristics group, is the most important overall, indicating that the length of the loan is a significant determinant of default risk. Longer loan terms might be more difficult for individuals to handle over time, increasing the risk of financial instability. Shortening loan terms or providing flexible repayment options may help borrowers manage their debt more successfully, lowering the probability of default. Jappelli and Pagano (2002) and Qi (2023) also found that longer loan periods increased the likelihood of default, emphasizing the need to carefully consider loan terms.

Active accounts number (11.62%): Within the credit history group, this feature has a high importance value. A larger number of open accounts is associated with an increased risk of default, probably due to the complexity and financial burden of handling many credit lines at the same time. Financial institutions should evaluate a borrower's total number of active accounts when determining creditworthiness. A larger number of open accounts may suggest overextension, and lenders should be careful before granting extra credit to such customers (Elhoseny et al., 2025).

Highest credit cap (10.88%): As the most important feature in the applicant financial information group, this indicates that the maximum credit limit plays a crucial role in determining a borrower's financial strength and potential risk of default. Higher credit limits are connected with a higher chance of default, most likely because people with high credit limits are more inclined to use credit, putting them under more financial hardship. Lenders may also explore establishing more conservative credit restrictions for customers with less reliable financial backgrounds. Gross and Souleles (2002) found that larger credit limits can lead to more borrowing and a higher chance of default, particularly among financially disadvantaged people.

Financing objective (9.51%): This feature from the loan characteristics group emphasizes that the purpose of the loan can significantly affect default risk. Loans for high-risk purposes might have higher default probabilities compared to those with stable and secure purposes. Different lending objectives may have distinct risk profiles. For example, loans for critical purposes such as house repair may be less hazardous than loans for frivolous expenditures. When analyzing credit risk, financial institutions should take into account the loan's purpose. This is corroborated by Avery and Samolyk's (2004) research, which shows that the purpose of the loan has a considerable influence on the chance of default.

Employment tenure (9.29%): This is the most critical feature in the employment and residential information group, underscoring the importance of job stability in assessing default risk. Longer job tenures are connected with a decreased probability of default, highlighting the relevance of work stability in financial health. Financial institutions might profit by emphasizing job stability during the loan application process. Applicants with longer job histories are more likely to have consistent earnings, which reduces the chance of default. Siddique et al. (2022) found that employment stability considerably decreases credit risk.

Residential status (9.07%): Also in the employment and residential information group, this feature shows that homeownership or stable residency may positively influence repayment capacity and reduce default risk. Homeowners are often regarded as more stable because of the equity in their house. Financial institutions should include house ownership as a major factor in loan decisions. Michalak and Uhde (2012) discovered that homeownership is a reliable predictor of financial stability and creditworthiness.

Outstanding credit (7.6%): This feature highlights the importance of existing debt levels in the applicant financial information group. Higher outstanding credit could increase the risk of default if it indicates over-leveraging. It may signal that a borrower is already experiencing financial hardship, making it more difficult to satisfy further financial commitments. To avoid defaults, financial institutions should regularly monitor credit balances and offer financial counseling or help to borrowers with large credit amounts. Borrowers can better manage their debt by taking proactive actions like debt consolidation or payback arrangements. This observation is supported by research from Kim et al. (2018), who found that high credit balances are strong indicators of financial distress and default risk.

Credit tenure (7.12%): Within the credit history group, the length of time a borrower has held credit accounts is important for predicting default, as longer credit histories may demonstrate creditworthiness. A well-managed, extended credit history is a good sign. During the credit rating process, financial institutions should take into account a borrower's credit history's length and quality. Avery et al. (2009) found that a lengthy and positive credit history is a substantial predictor of excellent credit performance.

Months since last default (7.1%): This feature from the credit history group indicates that recent defaults are a strong predictor of future default risk. More recent delinquencies indicate a larger chance of default, emphasizing the relevance of current credit behavior in risk evaluations. Financial institutions should rigorously scrutinize borrowers' recent credit histories. Dinh and Kleimeier (2007) found that recent delinquencies are a strong predictor of credit risk.

Principal amount (5.7%): Although relatively less important, the principal loan amount in the loan characteristics group still affects default risk, as larger loans might be harder to repay. Larger loan amounts enhance the borrower's financial stress, which directly correlates with the probability of default. Financial institutions should determine if the proposed loan amount is within the borrower's ability to repay. Riddiough and Wyatt (1994) found that bigger loan amounts are related to a higher risk of default, especially in non-collateralized loans.

Annual earnings (5.41%): This feature shows the significance of income levels in the applicant financial information group, with higher earnings generally reducing default risk. Annual earnings give additional resources for debt management, but they must be assessed in conjunction with total financial commitments and costs. Financial institutions should properly assess a borrower's annual earnings and compare them to their entire financial commitments to ensure repayment capability.

Monthly obligations (4.85%): The least important feature, monthly financial obligations still play a role in determining default risk, as high obligations can strain a borrower's repayment ability. Monthly obligations, which represent recurring monthly commitments such as rent or mortgage payments, have a substantial influence on a borrower's capacity to service extra debt. When assessing credit risk, financial institutions should carefully consider a borrower's monthly liabilities. According to Campbell and Cocco (2015), significant monthly obligations might have a detrimental impact on a borrower's capacity to repay debt.

The analysis of feature importance in the XGB model provides significant theoretical and practical contributions to credit risk management. Theoretically, the research findings further reinforce the critical role of key feature groups in predicting personal default risk, including applicant financial information, loan characteristics, and credit history. Among these, loan term, which belongs to the loan characteristics group, is identified as the most influential factor, as longer loan durations can increase financial pressure and the probability of default. Active accounts number, within the credit history group, is also an important feature that reflects credit utilization levels and a borrower's ability to manage financial obligations. Additionally, highest credit cap, from the applicant financial information group, plays a decisive role in risk assessment, as excessively high credit limits may lead to over-borrowing and repayment difficulties.

From a practical perspective, financial institutions need to regulate loan terms appropriately, monitor the number of active credit accounts, and flexibly adjust credit limits based on customers' financial capacity and repayment willingness. For instance, granting long-term loans should be accompanied by risk control measures such as setting reasonable loan term limits or applying flexible repayment plans, enabling borrowers to maintain their repayment ability without prolonged financial pressure. It is also essential to limit the number of credit accounts a borrower can open simultaneously and closely monitor customers with multiple active accounts but unstable cash flows. Furthermore, credit institutions should not only assess a customer's access to credit limits but also evaluate their actual credit utilization. A flexible credit policy that adjusts limits based on financial behavior rather than solely on income could help mitigate the risk of default.

5. CONCLUSION

This study demonstrates the effectiveness of various tree-based machine-learning models, such as DT, RF, and XGB, in predicting personal default risk. Among these, XGB consistently outperforms the other models across seven key evaluation metrics: confusion matrix, accuracy, precision, sensitivity, specificity, F1 score, and AUC. XGB also shows its ability to manage large, multidimensional financial datasets and capture complex nonlinear relationships between features. The gradient boosting mechanism of XGB enables the aggregation of weak classifiers into a strong predictive model, resulting in superior accuracy, robustness, and computational efficiency. These advantages are particularly crucial for financial institutions that require precise and scalable credit risk assessment models to optimize lending decisions and mitigate financial losses.

The feature importance analysis highlights three dominant predictors of default risk: loan term, active accounts number, and highest credit cap. Loan term stands out, emphasizing the significance of loan term duration in calculating default risk. The active accounts number shows the number of active credit accounts a borrower has, which might represent their credit management abilities and general financial behavior. A larger number of open accounts can indicate strong credit utilization, but it can also put the borrower in danger of becoming over-leveraged. Similarly, highest credit cap indicates the entire credit accessible to a borrower. a greater credit limit may signify strong creditworthiness, but it also raises the potential for excessive borrowing and financial strain.

Although this study makes significant contributions by providing empirical evidence on the effectiveness of machine learning models in predicting personal default risk, certain limitations remain that need to be addressed in future research. First, the study focuses solely on comparing the performance of basic tree-based machine learning algorithms and does not explore more complex approaches, such as hybrid models combining tree-based methods with deep learning. This limitation may restrict a deeper understanding of the effectiveness of tree-based algorithms across different methodological approaches. Second, while the study identifies the most important features influencing default risk prediction, it does not determine the direction of the impact of these features on the prediction outcome. Third, the findings are derived from analyzing personal financial data within the Vietnamese market, which may limit the generalizability of the results when applied to different economic and financial environments.

Therefore, to address these limitations, future research can explore the following directions. First, integrating tree-based models with deep learning algorithms should be considered to enhance the predictive performance of tree-based models in credit risk management. This can be achieved through models such as deep neural decision forests (DNDFs), neural-backed decision trees (NBDTs), or deep gradient boosting machines (DeepGBM). Additionally, to better understand the impact of feature variables on prediction outcomes, techniques such as SHAP (shapley additive explanations) and LIME (local interpretable model-agnostic explanations) should be utilized. Moreover, to regulate the relationship between feature variables and the target variable in the predictive model, monotonic constraints should be tested on the XGB model. Finally, to improve the generalizability of the study and ensure that the model can be applied across different markets, future research should consider testing on datasets from various financial markets or applying cross-validation across different markets.

REFERENCES

Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks, 6*(2), Article 38. https://doi.org/10.3390/risks6020038

Arifaj, A. H., & Baruti, B. H. (2023). The effect of credit risk on the financial performance of commercial banks in Balkan countries. *Corporate & Business Strategy Review, 4*(3), 18–25. https://doi.org/10.22495/cbsrv4i3art2

Avery, R. B., & Samolyk, K. A. (2004). Bank consolidation and small business lending: The role of community banks. *Journal of Financial Services Research, 25*, 291–325. https://doi.org/10.1023/B:FINA.0000020667.29802.1c

Avery, R. B., Brevoort, K. P., & Canner, G. B. (2009). Credit scoring and its effects on the availability and affordability of credit. *Journal of Consumer Affairs, 43*(3), 516–537. https://doi.org/10.1111/j.1745-6606.2009.01151.x

Avery, R. B., Calem, P. S., & Canner, G. B. (2004). Consumer credit scoring: Do situational circumstances matter? *Journal of Banking & Finance, 28*(4), 835–856. https://doi.org/10.1016/j.jbankfin.2003.10.009

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications, 83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Barua, S., Gavandi, D., Sangle, P., Shinde, L., & Ramteke, J. (2021). Swindle: Predicting the probability of loan defaults using CatBoost algorithm. In *Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1710–1715). IEEE. https://doi.org/10.1109/ICCMC51019.2021.9418277

Bao, W., Lianju, N., & Yue, K. (2019). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications, 128*, 301–315. https://doi.org/10.1016/j.eswa.2019.02.033

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Campbell, J. Y., & Cocco, J. F. (2015). A model of mortgage default. *The Journal of Finance, 70*(4), 1495–1554. https://doi.org/10.1111/jofi.12252

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). https://doi.org/10.1145/2939672.2939785

Deng, H., & Runger, G. (2013). Gene selection with guided regularized random forest. *Pattern Recognition, 46*(12), 3483–3489. https://doi.org/10.1016/j.patcog.2013.05.018

Dinh, T. H. T., & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis, 16*(5), 471–495. https://doi.org/10.1016/j.irfa.2007.06.001

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research, 297*(3), 1178–1192. https://doi.org/10.1016/j.ejor.2021.06.053

Elhoseny, M., Metawa, N., Sztano, G., & El-Hasnony, I. M. (2025). Deep learning-based model for financial distress prediction. *Annals of Operations Research, 345*, 885–907. https://doi.org/10.1007/s10479-022-04766-5

Eyalsalman, S., Alzubi, K., & Marashdeh, Z. (2024). The impact of IFRS 9, liquidity risk, credit risk, and capital on banks' performance [Special issue]. *Journal of Governance & Regulation, 13*(1), 396–404. https://doi.org/10.22495/jgrv13i1siart13

Fay, S., Hurst, E., & White, M. J. (2002). The household bankruptcy decision. *American Economic Review, 92*(3), 706–718. https://doi.org/10.1257/00028280260136327

Fitzpatrick, T., & Mues, C. (2021). How can lenders prosper? Comparing machine learning approaches to identify profitable peer-to-peer loan investments. *European Journal of Operational Research, 294*(2), 711–722. https://doi.org/10.1016/j.ejor.2021.01.047

Gao, J., Sun, W., & Sui, X. (2021). Research on default prediction for credit card users based on XGBoost-LSTM model. *Discrete Dynamics in Nature and Society,* Article 5080472. https://doi.org/10.1155/2021/5080472

Ghatasheh, N. (2014). Business analytics using random forest trees for credit risk prediction: A comparison study. *International Journal of Advanced Science and Technology, 72*, 19–30. https://doi.org/10.14257/ijast.2014.72.02

Goel, A., & Rastogi, S. (2023). Understanding the impact of borrowers' behavioural and psychological traits on credit default: Review and conceptual model. *Review of Behavioral Finance, 15*(2), 205–223. https://doi.org/10.1108/RBF-03-2021-0051

Gross, D. B., & Souleles, N. S. (2002). Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data. *The Quarterly Journal of Economics, 117*(1), 149–185. https://doi.org/10.1162/003355302753399472

Gumus, M., & Kiran, M. S. (2017). Crude oil price forecasting using XGBoost. In *Proceedings of the 2017 International Conference on Computer Science and Engineering (UBMK)* (pp. 1100–1103). IEEE. https://doi.org/10.1109/UBMK.2017.8093500

He, S., Zhou, L., Xie, H., & Tan, S. (2024). Enhancing XGBoost's accuracy in soil organic matter prediction through feature fusion. *Paddy and Water Environment, 22,* 475–489. https://doi.org/10.1007/s10333-024-00980-y

Hussin Adam Khatir, A. A., & Bee, M. (2022). Machine learning models and data-balancing techniques for credit scoring: What is the best combination? *Risks, 10*(9), Article 169. https://doi.org/10.3390/risks10090169
Iranzad, R., & Liu, X. (2024). A review of random forest-based feature selection methods for data science education and applications. *International Journal of Data Science and Analytics*. https://doi.org/10.1007/s41060-024-00509-w
Jappelli, T., & Pagano, M. (2002). Information sharing, lending and defaults: Cross-country evidence. *Journal of Banking & Finance, 26*(10), 2017–2045. https://doi.org/10.1016/S0378-4266(01)00185-6
Jiang, Y. (2022). Credit ratings, financial ratios, and equity risk: A decomposition analysis based on Moody's, Standard & Poor's, and Fitch's ratings. *Finance Research Letters, 46*, Article 102512. https://doi.org/10.1016/j.frl.2021.102512
Keese, M. (2012). Who feels constrained by high debt burdens? Subjective vs. objective measures of household debt. *Journal of Economic Psychology, 33*(1), 125–141. https://doi.org/10.1016/j.joep.2011.08.002
Kim, H., Cho, H., & Ryu, D. (2018). An empirical study on credit card loan delinquency. *Economic Systems, 42*(3), 437–449. https://doi.org/10.1016/j.ecosys.2017.11.003
Li, H., Cao, Y., Li, S., Zhao, J., & Sun, Y. (2020). XGBoost model and its application to personal credit evaluation. *IEEE Intelligent Systems, 35*(3), 52–61. https://doi.org/10.1109/MIS.2020.2972533
Li, Y., & Chen, W. (2020). A comparative performance assessment of ensemble learning for credit scoring. *Mathematics, 8*(10), Article 1756. https://doi.org/10.3390/math8101756
Liu, X., Wang, H., Zhang, K., Lin, K., Shi, Q., & Zeng, F. (2024). Credit default of P2P online loans based on logistic regression model under factor space theory risk prediction research. In Z. Shi, J. Torresen, & S. Yang (Eds.), *International Conference on Intelligent Information Processing* (Vol. 703: IFIP Advances in Information and Communication Technology, pp. 410–424). Springer. https://doi.org/10.1007/978-3-031-57808-3_30
Lu, Z., Li, H., & Wu, J. (2024). Exploring the impact of financial literacy on predicting credit default among farmers: An analysis using a hybrid machine learning model. *Borsa Istanbul Review, 24*(2), 352–362. https://doi.org/10.1016/j.bir.2024.01.006
Lee, C. S., Cheang, P. Y. S., & Moslehpour, M. (2022). Predictive analytics in business analytics: Decision tree. *Advances in Decision Sciences, 26*(1), 1–29. https://iads.site/wp-content/uploads/papers/2022/Predictive-Analytics-in-Business-Analytics-Decision-Tree.pdf
Michalak, T. C., & Uhde, A. (2012). Credit risk securitization and bank soundness in Europe. *The Quarterly Review of Economics and Finance, 52*(3), 272–285. https://doi.org/10.1016/j.qref.2012.04.008
Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications, 161*, Article 113567. https://doi.org/10.1016/j.eswa.2020.113567
Naili, M., & Lahrichi, Y. (2022). The determinants of banks' credit risk: Review of the literature and future research agenda. *International Journal of Finance & Economics, 27*(1), 334–360. https://doi.org/10.1002/ijfe.2156
Nguyen, N., & Ngo, D. (2025). Comparative analysis of boosting algorithms for predicting personal default. *Cogent Economics & Finance, 13*(1). Article 2465971. https://doi.org/10.1080/23322039.2025.2465971
Pan, P.-J., Lee, C.-H., Hsu, N.-W., & Sun, T.-L. (2024). Combining principal component analysis and logistic regression for multifactorial fall risk prediction among community-dwelling older adults. *Geriatric Nursing, 57*, 208–216. https://doi.org/10.1016/j.gerinurse.2024.04.021
Pate, A., Riley, R. D., Collins, G. S., Van Smeden, M., Van Calster, B., Ensor, J., & Martin, G. P. (2023). Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. *Statistical Methods in Medical Research, 32*(3), 555–571. https://doi.org/10.1177/09622802231151220
Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3–14. https://doi.org/10.1080/00220670209598786
Qi, X. (2023). Factors influence loan default — A credit risk analysis. In X. Li, C. Yuan, & J. Kent (Eds.) *Proceedings of the 7th International Conference on Economic Management and Green Development* (Applied Economics and Policy Studies, pp. 849–862). Springer. https://doi.org/10.1007/978-981-97-0523-8_79
Riddiough, T. J., & Wyatt, S. B. (1994). Strategic default, workout, and commercial mortgage valuation. *The Journal of Real Estate Finance and Economics, 9*(1), 5–22. https://doi.org/10.1007/BF01153586
Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *WIREs: Data Mining and Knowledge Discovery, 8*(4), Article e1249. https://doi.org/10.1002/widm.1249
Shetty, S., Musa, M., & Brédart, X. (2022). Bankruptcy prediction using machine learning techniques. *Journal of Risk and Financial Management, 15*(1), Article 35. https://doi.org/10.3390/jrfm15010035
Siddique, A., Khan, M. A., & Khan, Z. (2022). The effect of credit risk management and bank-specific factors on the financial performance of the South Asian commercial banks. *Asian Journal of Accounting Research, 7*(2), 182–194. https://doi.org/10.1108/AJAR-08-2020-0071
Suhadolnik, N., Ueyama, J., & Da Silva, S. (2023). Machine learning for enhanced credit risk assessment: An empirical approach. *Journal of Risk and Financial Management, 16*(12), Article 496. https://doi.org/10.3390/jrfm16120496
Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications, 237*, Article 121549. https://doi.org/10.1016/j.eswa.2023.121549
Uddin, M. S., & Rahman, M. A. (2024). A comparative study of machine learning algorithms for enhanced credit default prediction. In D. Asirvatham, F. M. Gonzalez-Longatt, P. Falkowski-Gilski, & R. Kanthavel (Eds.), *Evolutionary Artificial Intelligence* (Algorithms for Intelligent Systems, pp. 201–216). Springer. https://doi.org/10.1007/978-981-99-8438-1_15
Velarde, G., Weichert, M., Deshmunkh, A., Deshmane, S., Sudhir, A., Sharma, K., & Joshi, V. (2024). Tree boosting methods for balanced and imbalanced classification and their robustness over time in risk assessment. *Intelligent Systems with Applications, 22*, Article 200354. https://doi.org/10.1016/j.iswa.2024.200354
Ying, L. (2018). Research on bank credit default prediction based on data mining algorithm. *International Journal of Social Sciences and Humanities Invention, 5*(6), 4820–4823. https://doi.org/10.18535/ijsshi/v5i6.09
Zheng, H., Yuan, J., & Chen, L. (2017). Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies, 10*(8), Article 1168. https://doi.org/10.3390/en10081168
Zhou, Y. (2023). Loan default prediction based on machine learning methods. In *Proceedings of the 3rd International Conference on Big Data Economy and Information Management*. https://doi.org/10.4108/eai.2-12-2022.2328740