FMDB *Transactions on Sustainable Finance and Data Science*

A Comparative Analysis of Supervised Machine Learning Techniques to Predict Loan Defaults in Data-Limited Contexts

Rejwan Bin Sulaiman[1,*], Sahiti Puppala[2]

[1,2]Department of Computer and Information Sciences, Northumbria University, London, England, United Kingdom.
rejwan.sulaiman@northumbria.ac.uk[1], sahiti.puppala@gmail.com[2]

Abstract: The paper has never been more important to accurately forecast borrower behaviour, where a lending decision could risk financial stability. Although the potential of credit assessment models is being explored, they are confined to traditional financial information or a narrower territorial scope. By using data samples from two different lending platforms, where borrowers' credit histories are frequently unclear and have a broad geographic distribution, this study offers a more comprehensive approach to credit risk analysis. For the investigation, five ML models, namely Logistic Regression, Decision Tree, Random Forest, Gradient Boosting and XGBoost, are employed. The models are trained and tested using a combination of sourced samples and evaluated using metrics such as accuracy, precision, recall, F1 score, confusion matrix and ROC AUC. The findings indicate that ensemble models, particularly XGBoost, are most effective, with consistent performance across all metrics. It is also identified that characteristics such as loan amount, interest rate, instalment amount, term period and demographics such as annual income, gender, home ownership and country of origin influence repayment behaviour. These insights contribute to the field of credit risk assessment with practical implications of reducing default rates that improve lenders' profitability and stability.

Keywords: Decision Tree; Repayment Behaviour; Lending Platform; Financial Stability; Gradient Boosting; Random Forest; Logistic Regression; Credit Risk Assessment; Instalment Amount.

Received on: 04/06/2025, **Revised on:** 27/07/2025, **Accepted on:** 12/10/2025, **Published on:** 07/03/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSFDS

DOI: https://doi.org/10.69888/FTSFDS.2026.000622

Cite as: R. B. Sulaiman and S. Puppala, "A Comparative Analysis of Supervised Machine Learning Techniques to Predict Loan Defaults in Data-Limited Contexts," *FMDB Transactions on Sustainable Finance and Data Science*, vol. 1, no. 1, pp. 19–38, 2026.

1. Introduction

In the dynamic financial sector, where precision is essential, credit risk assessment has become fundamental, especially for organisations serving vulnerable populations [15]. The ability to estimate credit risk is more critical when it directly affects an institution's ability to manage defaults and preserve financial equilibrium. In this context, technological advancements in ML are revolutionising credit risk assessment [8]. Credit assessment is a methodical, scientific evaluation of a borrower's creditworthiness, a vital tool in financial services. Traditionally, a numerical score is calculated to reflect the borrowers' likelihood of repaying a loan [1]. According to Noriega et al. [36], this straightforward score provides an impartial risk

[*]Corresponding author.

assessment and accelerates decision-making. However, this evaluation has its own limitations, as it relies heavily on historical customer data, which disadvantages those with low scores or limited credit records [31]. With the application of ML techniques, a revolution has occurred, enabling the analysis of a wide range of borrower data to forecast repayment behaviour with high precision. As previous research demonstrates, these models predict risk more accurately. However, the question remains unanswered whether these models may withstand the challenge of insufficient data [4]; [34]. Many clients have sparse credit histories that traditional scoring models struggle with Paraíso et al. [37]. The lack of borrower history is also common in organisations such as microfinance institutions, where clients often lack access to traditional financial services [4]. Some researchers have attempted to bridge this void by combining other data sources, including email, social network contacts, utility bill payments and so on Cornée [10] and Li et al. [27]. While these methods have demonstrated efficacy, they raise new concerns about user privacy [43]. Considering these complications, forecasting a borrower's payment behaviour solely on readily accessible information remains uncertain [22]. To address this question, this study uses a combination of two data sets from different online lending platforms. One sample represents the extensive data available for a borrower, while the other has basic information about clients from various countries. Furthermore, several studies used data from organizations in a particular region, limiting their findings to those locations [20].

For instance, Durango-Gutiérrez et al. [14] focused on organisations in Bolivia and Colombia, whereas Condori-Alejo et al. [9] concentrated on Persian institutions. Therefore, to offer diverse insights and confirm the conclusions of previous researchers, the following subsection outlines the aims of this study. This study aims to determine factors that affect credit risk and explore the feasibility of using ML to forecast borrower behaviour when traditional credit history is unavailable. The main objectives of this study are to evaluate the effectiveness of various ML models in predicting loan performance using minimal, readily accessible data; to perform cross-regional analysis using data from online lending platforms; and to identify factors that influence default probability. The significance of this research lies in both theoretical advancements and practical applications. This research confirms existing theories in credit risk assessment and provides insights for financial organisations. Organisations seeking to understand borrowers' financial practices across different regions can benefit from this study. The conclusions might also help enhance credit evaluation protocols for organizations. However, this study does not discuss qualitative aspects, such as the influence of culture on credit risk assessment. This study is also limited in its exploration of the integration of non-financial sources for users. This study confines itself to evaluating the standard model’s performance in the absence of generally available information. This research does not include the proposal or exploration of new methods for credit risk assessment. Ultimately, this study utilizes secondary data; therefore, the conclusions may apply to organizations similar to the source organization.

2. Literature Review

Conventional banks have served as the primary source of consumer loans over the years [16]. Usually, lending money to households with substantial assets, high levels of education and stable incomes [33]. For many borrowers, however, obtaining loans through traditional banks presents challenges [3]. Particularly those with limited financial history, uncommon profiles and underserved communities [39]; [6]. These barriers to financial inclusion have led to the adoption of digital or contemporary financial services as alternatives. Since traditional financial institutions discriminate against certain groups, many customers prefer modern loan services despite high interest rates. Services such as peer-to-peer (P2P) lending and microfinance institutions (MFIs) have become major competitors in the market [49]. However, these platforms often struggle to balance their growth with lenders’ inclination toward low-risk borrowers [24]. Without verifiable financial history or sources, these platforms tend to attract riskier borrowers, which contributes to high default rates. In fact, modern lending systems have up to 90% of their customers on the blocklist [21]. These platforms are increasingly implementing ML techniques to address these issues. When forecasting credit risk, ML models have outperformed statistical methods in terms of accuracy [6]; [41]. The regulatory capital requirements have also decreased as a result of their integration falling from 17% to 12.4% [2]. However, there is still debate over the application of ML to credit risk assessment. Critics highlight concerns about fairness, transparency, and accuracy despite their persistence in human decision-making [5]. In the literature, addressing these concerns is still a crucial area of growth.

2.1. Machine Learning Approaches to Credit Scoring

Research in credit risk assessment using ML has advanced significantly. Researchers have also attempted to improve accuracy by combining algorithms. However, their implementation is still limited, especially when borrower data is unclear [41].

2.1.1. Logistic Regression

The ease of use and interpretability of logistic regression have made it a staple choice in risk assessment literature. Even with its reliable performance, Logistic Regression frequently performs similarly to advanced models. Studies by Condori-Alejo et al. [9] and Pradnyana [38], for example, have reported that Logistic Regression has high accuracy in predicting default probability. However, it is often outperformed by ensemble methods. Medina-Olivares et al. [32] have also highlighted the efficiency of Logistic Regression and demonstrated its interpretability by identifying significant factors. These findings suggest that although Logistic Regression is useful, it is sometimes outperformed by advanced algorithms.

2.1.2. Support Vector Machine

The use of support vector machines in credit risk assessment has produced inconsistent results despite their effectiveness in high-dimensional spaces. Jakka et al. [19] reported superior performance when SVMs were combined with Gain Ratio feature selection, though this was also dataset-specific. However, Condori-Alejo et al. [9] and Lusinga et al. [30] implemented SVMs but found that other models outperformed them. Nalić et al. [34] note that SVMs are computationally demanding and underperforming, leading to their exclusion from the final analysis. These differences imply that SVM’s efficiency varies highly depending on the context.

2.1.3. K- Nearest Neighbours

KNN is a non-parametric model; however, its performance in credit assessment research has been inconsistent. Condori-Alejo et al. [9] and Pradnyana [38] found that KNN performed well but was outrun by advanced models. In contrast, Ampountolas et al. [4] reported modest performance for the KNN algorithm and noted that its difficulties increased with the number of features. In this context, it is worth noting that Li [28] demonstrated that feature selection methods can enhance KNN’s performance. Drawing from these conclusions, KNN might not be the optimal choice, but it has potential for improvement.

2.1.4. Decision Trees

Decision trees are usually effective at handling non-linear data, but researchers have identified limitations in credit risk assessment. Dumitrescu et al. [13] note that decision trees are difficult to interpret and are prone to overfitting. The researcher later proposes a hybrid model of a decision tree and logistic regression to address this issue. Klusowski and Tian [25] also highlight this model’s tendency to overfit and its impact in large-scale production. These limitations should be considered when implementing this model for credit risk evaluation.

2.1.5. Ensemble Methods

Ensemble methods have high accuracy and can capture linear and non-linear data. In previous research, Suhadolnik et al. [44] noted that these models perform better than a single algorithm. According to Lusinga et al. [30] and Pradnyana [38], ensemble models, specifically XGBoost, performed better than traditional models. However, it is worth noting that amongst the ensemble models, Gradient Boosting is constantly outperformed by Random Forest and XGBoost. These findings are further supported by additional research [18]; [46]; [42].

2.1.6. Artificial Neural Networks

As ANNs can represent intricate and non-linear relations, they also have potential for use in credit rating. Numerous researchers have demonstrated the efficacy of ANNs, including Condori-Alejo et al. [9] and Durango-Gutiérrez et al. [14]. These researchers have concluded that their algorithms achieve higher accuracy than their base algorithms. However, some researchers have identified that ensemble methods perform better than ANNs. They include Jakka et al. [19] and Shukla et al. [42]. Although ANNs are powerful, the contradicting results imply their performance might be context-dependent.

2.2. Common Factors Influencing Default Risk

Various features were observed to impact credit risk, including loan characteristics, borrower demographics and transactional data. This section reviews the commonly found significant factors.

2.2.1. Loan Characteristics

Financial data, particularly loan-related information, is critical for predicting creditworthiness. Previous research indicates that larger loan amounts and higher interest rates generally correlate with default risk [4]. Ampountolas et al. [4] and Medina-Olivares et al. [32] found that shorter loan terms were associated with lower risk, suggesting that borrowers were less likely to default. Guarantor information has also impacted the probability of default [14]. However, these researchers note the significance of exploring other factors beyond loan characteristics.

2.2.2. Borrowers' Demographics

Demographic features such as age, gender, ethnicity, educational level, and marital status were frequently examined. Studies emphasise the influence of these attributes on financial behaviour [4]; [9]; [38]. Employment status was another factor observed to influence repayment [44]. Although demographic information is relevant, it is important to approach it with caution, as it can introduce bias into the predictions. Researchers like Khatir and Bee [18] and Kozodoi et al. [26] have highlighted the importance of evaluating models across diverse populations.

2.2.3. Transactional Data

Transaction behaviour plays a major role in assessing credit risk and provides insights into borrowers' practices. Some strong predictors include payment history, repayment patterns, and overall financial conduct, as noted by Condori-Alejo et al. [9] and Durango-Gutiérrez et al. [14]. However, arguments about privacy and ethical concerns suggest a refinement of its integration.

2.2.4. Other Information

Interestingly, macroenvironmental variables were observed to have a great impact on credit risk. According to studies by Lusinga et al. [30] and Medina-Olivares et al. [32], borrowers may present a higher risk if they are located in economically deprived or unstable regions. Factors such as local unemployment rates and climatic conditions that affected livelihoods were analyzed. Medina-Olivares et al. [32] note that although these variables improve accuracy, they introduce complexity that may affect the model's reliability.

2.3. Addressing Class Imbalance

The fairness of credit scoring systems is seriously threatened by bias in ML algorithms. Several approaches have been proposed to address this problem, including sampling techniques. Several studies have used sampling techniques to address class imbalance, which is a common source of bias in ML models. Ampountolas et al. [4] used the Synthetic Minority Oversampling Technique (SMOTE) to balance nominal and continuous variables. On the other hand, Khatir and Bee [18] explored multiple sampling techniques and found that random sampling with random forests yielded the best results. Another researcher, Li [28], has shown the effectiveness of the Adaptive Synthetic (ADASYN) sampling method. Other researchers, such as Suhadolnik et al. [44], have utilised subsampling methods to address class imbalance. These diverse methods stress the importance of selecting appropriate sampling methods in credit scoring applications [45]. This literature review analyses various studies that have implemented ML in credit scoring.

The synthesis of existing literature showed advancements and challenges in their implementation as well. While traditional models are widely used, advanced techniques like ensemble methods have demonstrated better performance. This was evident in studies by Lusinga et al. [30], Pradnyana [38], Tumuluru et al. [46], Shukla et al. [42], Suhadolnik et al. [44], and Nalić et al. [35]. This suggests that future research could benefit from exploring advanced models or combining their strengths to develop hybrid models to address the growing challenges in lending. Secondly, the factors influencing the credit rate consist of a wide range of features. Studies found that demographic, transactional, macro-environmental, and guarantor data were crucial for assessing credit risk. However, these researchers have also suggested a careful integration of different information, as it may introduce bias. Lastly, the class imbalance problem, which introduces bias in model predictions, was explored using various sampling techniques. These techniques range from advanced sampling methods to adversarial biasing [4]; [18]; [28]; [44].

2.3.1. Summary Table

The following Table 1 summarises the methods and findings of key papers that align closely with this research.

Table 1: Summary table of key research papers

Reference	Data	Models	Methodology	Result	Suggestions
Ampountolas et al. [4]	Obrained from Innovative Microfinance Limited	Decision Trees, Extra Tree Classifier, Random Forest,	Removed 46 entries due to missing values.	The top three features are age, loan amount, and annualised rate.	Future studies focus on fair lending rates in the microcredit context.

	Contains loans from 2012 to 2018 Sample size of 4450 Includes demographic information, loan amount, frequency, outstanding balance and number of days in arrears.	XGBoost, AdaBoost, KNN, MLP	Transformed data for the loan amount using logarithmic transformations. Applied the SMOTENC sampling technique. 80/20 split Performance metrics include accuracy, precision, recall, F1 score, confusion matrix and AUC. Permutation importance score for feature importance.	The top three best performing models are Random Forest, XGBoost, and Adaboost, with each above 80%	
Condori-Alejo et al. [9]	Obtained from the microfinance institution database March 2017 to March 2018 Total records are 17454 from 15015 clients.	Logistic Regression, Random Forest, SVM, ANN, Decision tree, KNN	Used one-hot encoding for categorical variables. 75/25 split Performance metrics include accuracy, precision, recall, F1 score, and AUC.	The best model is the ANN with	Future work to expand results and analyze them outside the Persian region to validate them.
Durango-Gutiérrez et al. [14]	Collected data from 2 MFIs in Colombia and Bolivia .	Logistic Regression, ANN- MLP	10-fold cross-validation	Significant positive factors - Total arrears, average arrears debt ratio, loan amount, interest rate, guarantees and credit analysts' forecast for Columbia	Female gender and the number of previously granted loans for Bolivia. MLP performed better with 98.41% and 96.56%.

	2012-2015		75/25 split	Number of payments in default, liquidity ratio, loan amount, guarantees and credit analysts' forecast for Bolivia	
	4758 entries in total for Bolivia		Performance metrics used include accuracy and AUC	Significant negative factors – Female gender, MFI client history, Liquidity ratio, ROE, and COLCAP for Columbia	
	2627 entries for Columbia.			Female gender and the number of previously granted loans for Bolivia.	
				MLP performed better with 98.41% and 96.56%.	
Medina-Olivares et al. [32]	Obtained from an MFI based in China	Logistic regression with random effects.	Uses only independent random effects in logistic regression, spatial random effects using conditional autoregressive models	Loan size, ethnic minority status, local unemployment rate and education level are positively associated with group loan default.	Suggests focusing on the macro-environment to increase social sustainability.
	8513 individual loans and 15348 group loans in the sample.		To measure the importance of spatial dependence, a mixture between a model with spatial random effects and one with independent random effects	Also, identified group loans make borrowers' characteristics less important.	
	2017-2018		The integrated nested Laplace approximations are computationally intensive, as the models are.		

	Includes details about macroeconomic indicators from the China Provisional Statistical Yearbook.		Metrics are AUC, h-index, gini, KS statistic and logscore		
Pradnyana [38]	Data collected from multiple MFIs includes demographics, Loan characteristics, repayment histories and group-level variables. contains 161715 entries.	XGBoost, Logistic Regression, Decision Tree, Random Forest, KNN, and Linear Discriminant Analysis	Used AUC, Accuracy, Precision and Recall. Adjusted the weights of the majority and minority classes in the training phase.	XGBoost is the best model with 97% accuracy.	None specified

3. Methodology

3.1. Research Design

Following the framework proposed by Saunders [40], this research adopted a positivist perspective, focusing solely on observable, quantitative aspects. This philosophy was appropriate because it emphasises quantitative data, statistical analysis, and hypothesis testing, all of which are required for this research. Although pragmatism can offer flexibility and practicality, it can compromise the generalizability of the results that the study prioritises. Moreover, positivism provides a structured, objective approach that closely aligns with the goals of this research. Furthermore, this method can provide an impartial assessment that can guide the lender's judgment. Saunders [40] also categorised research approaches into deductive, inductive or a combination of both. This research uses a deductive approach, focusing on testing and validating pre-existing conclusions. These conclusions include previously identified features or a basic understanding of model performance. The goal was to validate these theories rather than develop new ones. While the inductive approach is intended to yield new theories or advances in the field, it does not align with the nature of this study or its objectives. Therefore, on this basis, the research primarily adopted a deductive approach rather than an inductive approach. Furthermore, the main objective of this study was to compare several algorithms (specifically, five) and test their performance. This led the research to lean toward an experimental strategy. This choice also facilitates feature exploration in this analysis, allowing identification of factors. This strategy also aligns with hypothesis testing and the measurement of dependent variables, which, in this case, is the default probability.

3.2. Data Collection

To evaluate the model's performance in varying data conditions, two datasets were combined to train and test the models. This approach allows us to assess the algorithm's effectiveness in two distinct lending decision contexts. The first one involves a comprehensive client profile, while the other involves limited information. This dual-dataset approach was inspired by previous studies by Medina-Olivares et al. [32] and Durango-Gutiérrez et al. [14]. The two samples used in this research are from Lending Club and Kiva. Combining these two data samples seemed to be a strategic approach for several reasons. It represents real decision-making scenarios and improves the model's resilience, flexibility, and generalizability. As a result, risk evaluation becomes more reliable and accurate, increasing the model's applicability in diverse environments.

3.2.1. Data Preparation

The datasets may contain missing values and inconsistent data types; as active loans cannot accurately reflect loan performance. They were excluded from the analysis to ensure the model was trained on factual records. The major pre-processing steps performed before analysis are listed as follows:

- **Handling Missing Values:** Following the suggestions of Zahid et al. [50] and Caton et al. [7], this study removed columns with more than 80% missing data to avoid imputation bias. This reduced the columns from 151 to 84. The remaining missing values were filled with placeholders (e.g., unknown for categorical variables and 0 for numerical variabl*es*). These records were further flagged to indicate that they were imputed into the model.

- **Encoding:** To efficiently encode categorical variables while minimising the dimensionality, label encoding was chosen over a one-hot encoder. This method captures relationships while maintaining a manageable feature space.
- **Feature Scaling:** The dataset's different statistical properties required a consistent transformation. Since models like Logistic Regression, Gradient Boosting, and Extreme Gradient Boosting that depend on distance were implemented in this study, feature scaling was crucial. Furthermore, the dataset consists of outliers that represent a diverse population. A StandardScaler was implemented because it is less sensitive to outliers than a MinMaxScaler.
- Additionally, the dataset underwent several other preprocessing steps. These include verification of duplicates, identifying and handling outliers, and consolidating similar categories. For example, entries such as 'DebtC' and 'Debt Consolidation' in the title column were merged to ensure consistency in data. This is followed by splitting the data into 80/20 for model development.
- The dataset consisted of 21.03% defaulted loans and 78.97% completed loans. As such, a class imbalance could introduce bias in the predictions, so the Random Under-Sampling Technique was used to address it. While SMOTE (Synthetic Minority Oversampling Technique) was also considered, it required high computational rigour. Since SMOTE increases the dataset size, it leads to higher computational demands. To manage resources effectively, random under-sampling was used to reduce the majority class size and balance the dataset.

3.3. Data Analysis

Statistical tests were also employed in this research to validate the findings of the data analysis, which identified key factors influencing credit risk. Thus, the following null hypotheses were formulated:

- **H_{01}:** There is no association between loan characteristics such as loan amount, interest rate, and instalments with loan default outcomes.
- **H_{02}:** There is no relationship between borrower demographics, including gender, employment sector, annual income, country of origin, and home ownership, with their repayment behaviour.

Rationale for t-Tests: To support the results for numerical variables, a t-test was employed. This test compares the means of two groups (e.g., class 0 vs class 1) and indicates whether any difference is statistically significant. It was chosen for its sensitivity in detecting minute differences in means. This is suitable for numerical features such as loan characteristics (loan amount, instalments, interest rate) and borrowers' annual income. However, this test assumes the dataset is normally distributed. Given the large dataset, this assumption might not be a significant concern.

Rationale for Chi-Square Tests: The chi-square test was chosen to validate the observations of categorical variables. This test assesses the degree of correlation between the samples' categorical variables. It is particularly suitable for borrower demographics (gender, sector, country, home ownership) and term period. This test does not assume data distribution and determines whether observed differences are statistically significant. This test also handles large datasets and provides reliable results.

3.4. Model Development

Supervised models were chosen due to the dataset's nature, which consisted of mapped input features to predetermined outcomes. Based on their unique benefits in handling various aspects, this study implemented logistic regression, decision trees, random forests, gradient boosting and XGBoost:

- Since understanding the relationship between dependent and independent variables is the main objective, logistic regression was selected for its interpretability and simplicity [12].
- Decision trees provide an intuitive representation, making them ideal for straightforward tasks. However, these models are prone to overfitting [11].
- Random Forests are implemented to mitigate overfitting and leverage ensemble learning, thereby improving predictive accuracy [17].
- As the dataset employed in this study is complicated, gradient boosting was selected, which is particularly effective for imbalanced data.
- Ultimately, XGBoost was chosen for enhanced computational efficiency and its proven superior performance in the literature. This model is beneficial for tasks that require high accuracy and speed [29].

This choice utilises the strengths of each algorithm and provides insights for credit risk assessment modelling.

3.5. Model Evaluation

The model's judgment and preference depend on the metric that aligns with the organisation. Some of the common metrics in ML applications for credit risk assessment are:

- **Confusion Matrix:** An evaluation metric for classifiers. It is a matrix with four combinations of predicted and actual labels. These combinations are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). This metric provides a detailed view of the model's performance across different prediction types.
- **Accuracy:** The simplest evaluation metric. It is calculated by dividing the correct number of predictions by the total number of predictions, as shown below. It is not ideal for imbalanced data; as high accuracy can be misleading if the model predicts only the majority class correctly:

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

- **Precision:** It is the measurement of the accuracy of positive predictions. Thus, it is defined as the ratio of true positives to the total number of positive observations. It can be crucial for credit risk management; as false positives can be expensive. For instance, lenders who want to avoid mistaking credit risk in applications that would repay:

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

- **Recall:** the ratio of true positives to the total number of actual positives in the data. It is often used along with precision. This metric measures the model's ability to correctly predict positive classes. For example, high recall flags most defaulters and reduces the probability of granting loans to risky applications:

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

- **F1 Score:** It is the harmonic mean of precision and recall. This metric is useful when both precision and recall are important. This score can be useful when a balance between precision and recall is required. A high score indicates the model's capability to avoid false positives and false negatives:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

- **Receiver Operating Characteristic (ROC) Curve:** It is a visual representation of a classification model's performance across all classification thresholds. It plots the true positive rate and false positive rate. This metric can help the lender visualise the model's efficiency between defaulters and non-defaulters.
- **Area Under the Curve (AUC):** It gives an aggregate measure of performance across all classification thresholds. However, it ignores the inhomogeneity of the ROC curve.

As the application requirements vary, this study has chosen to assess the models across all these metrics. Furthermore, using all metrics provides a comprehensive understanding of the model's performance. Therefore, the models will be evaluated using AUC-ROC, Recall, F1 Score and accuracy and precision.

3.6. Tools and Materials

Using Python libraries like Pandas, Numpy, Matplotlib, Seaborn and Scikit-learn, the analysis for this study was carried out. The machine used for this research is equipped with an Intel(R) Core(TM) i7-10510U CPU @ 1.80 GHz (2.30 GHz) and 8 GB RAM. Along with its 64-bit architecture and Google Colab for processing power, the configuration was ideal for data processing, machine learning and visualisation tasks.

3.7. Limitations

While the previous sections argued that the chosen methods were the best approach, the research is observed to have several methodological limitations:

- The positivist perspective emphasises objectivity, overlooking subjective insights and qualitative aspects, limiting the study's ability to understand contextual factors.
- The study used secondary datasets due to the unavailability of real-world data, which compromised the reliability of the models employed.
- Computational limitations prevented testing larger datasets or more advanced techniques, restricting the comprehensiveness of the results.

4. Results and Analysis

This section presents findings from evaluating various machine learning models and key factors that influence credit risk. The results are organized into a summary of data analysis and distribution, followed by a discussion of model development and evaluation.

4.1. Data Distribution and Analysis

This section explores the dataset's distribution across key variables and the insights derived from borrower demographics, loan characteristics, and other critical factors that impact loan status.

4.1.1. Geographic and Country-wise Loan Distribution

The dataset spans North America, South America, Africa, Asia, and parts of Europe, providing a diverse range of borrower demographics. The United States has the largest representation (600,000 entries), followed by the Philippines (180,000), Kenya (90,000), and several other countries with 10,000-40,000 entries each. This spread enables a comparative analysis of credit risk across regions. However, the distribution imbalance suggests caution in generalising the findings (Figure 1).

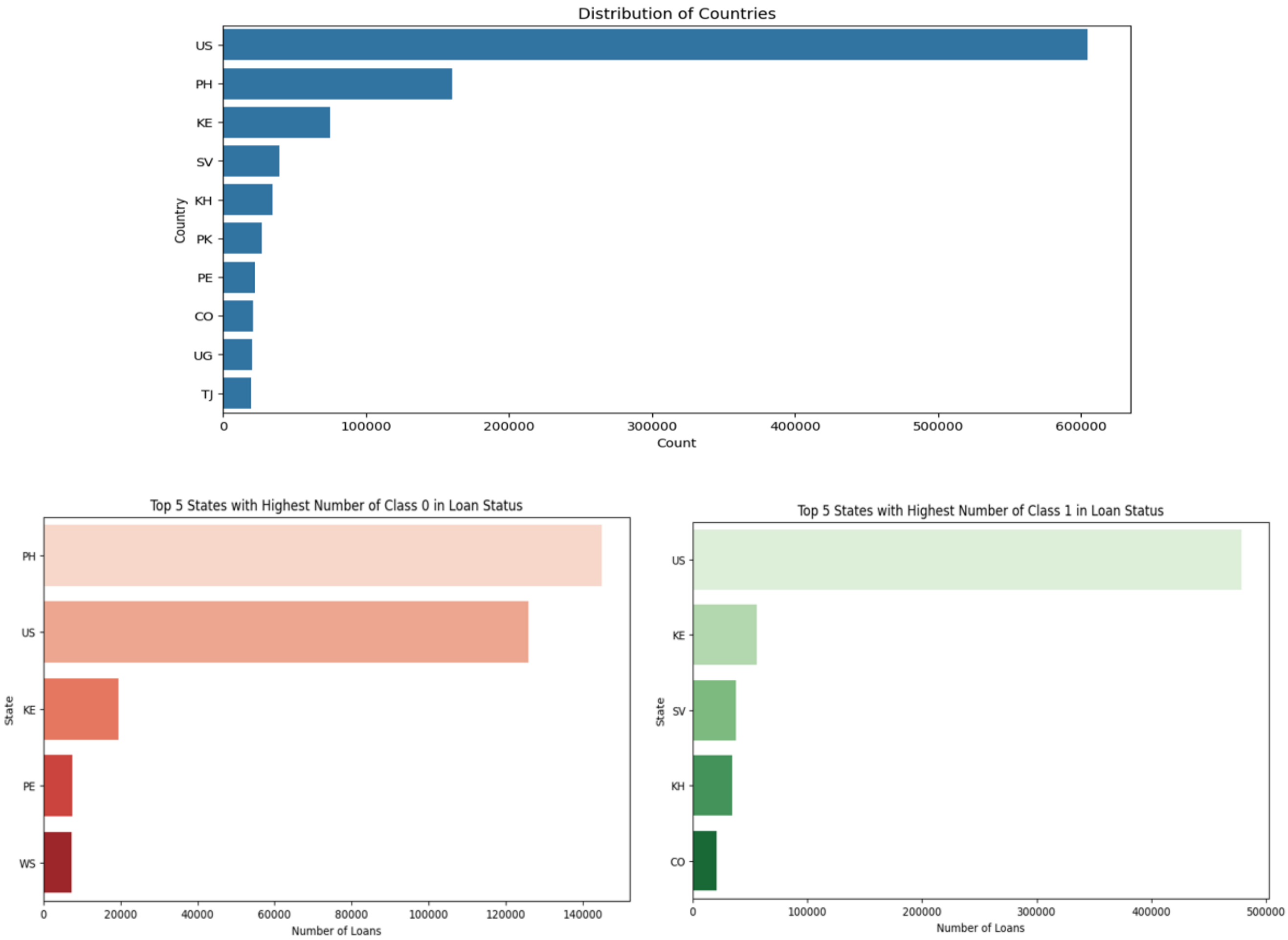

Figure 1: Geographic distribution of the dataset

4.1.2. Gender Distribution and Loan Status

The dataset's gender distribution is skewed, with 77.9% female borrowers and 22.1% male borrowers. Analysis shows that female borrowers have a higher repayment rate, with only about 40% of their loans defaulting, compared with a default rate below 20% among male borrowers. This skewness implies gender-specific borrowing trends, potentially influencing model predictions (Figure 2).

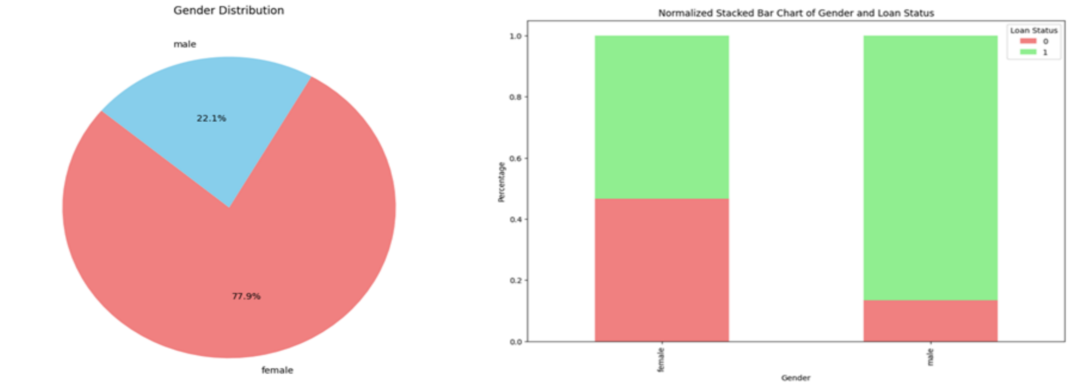

Figure 2: Gender distribution of the dataset

4.1.3. Loan Amount and Annual Income Distribution

Loan amounts are primarily skewed toward lower values, with most loans below 20,000 and a few outliers extending up to 100,000. Similarly, annual income is skewed toward the lower end, with a few high-income outliers. Borrowers with lower incomes are generally more likely to default on loans. The median income of non-defaulters is notably higher, with a marked distributional shift between defaulters and non-defaulters (Figure 3).

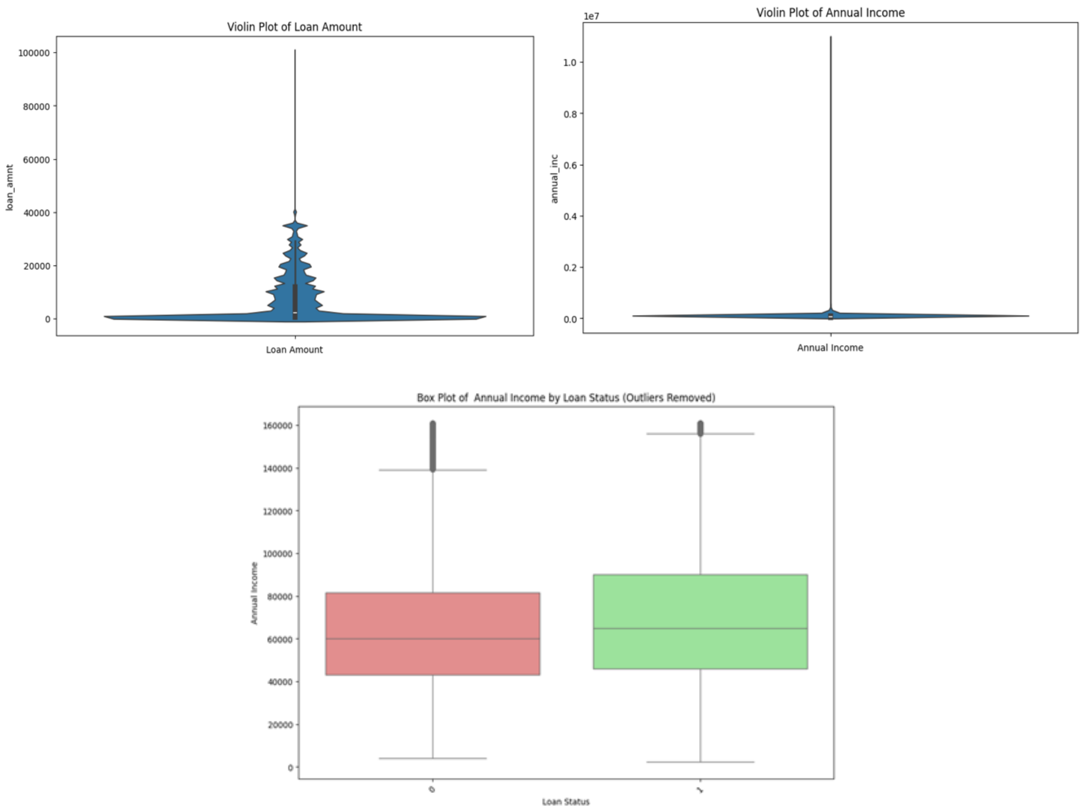

Figure 3: Loan amount and annual income distribution

4.1.4. Home Ownership

Homeownership status correlates with loan default rates: borrowers in the "rent" category have the highest default rates. In contrast, those in the "none" and "any" categories show the lowest (Figure 4).

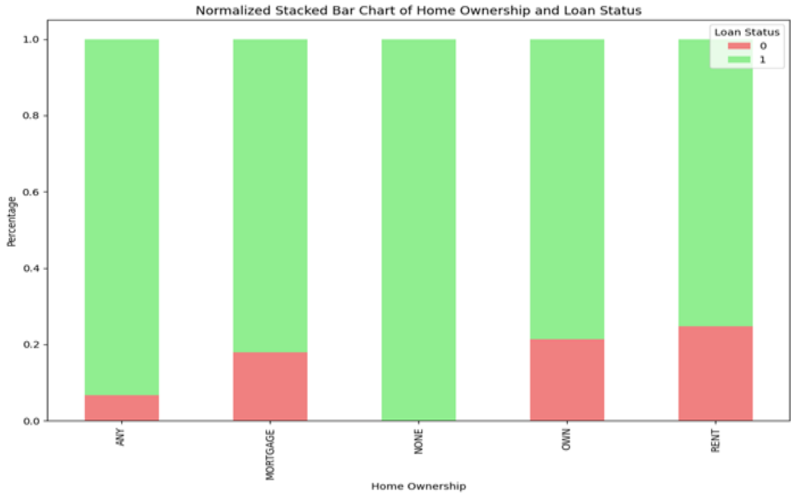

Figure 4: Home ownership distribution

4.1.5. Loan Characteristics: Term Period, Loan Amount, Interest Rate, and Instalments

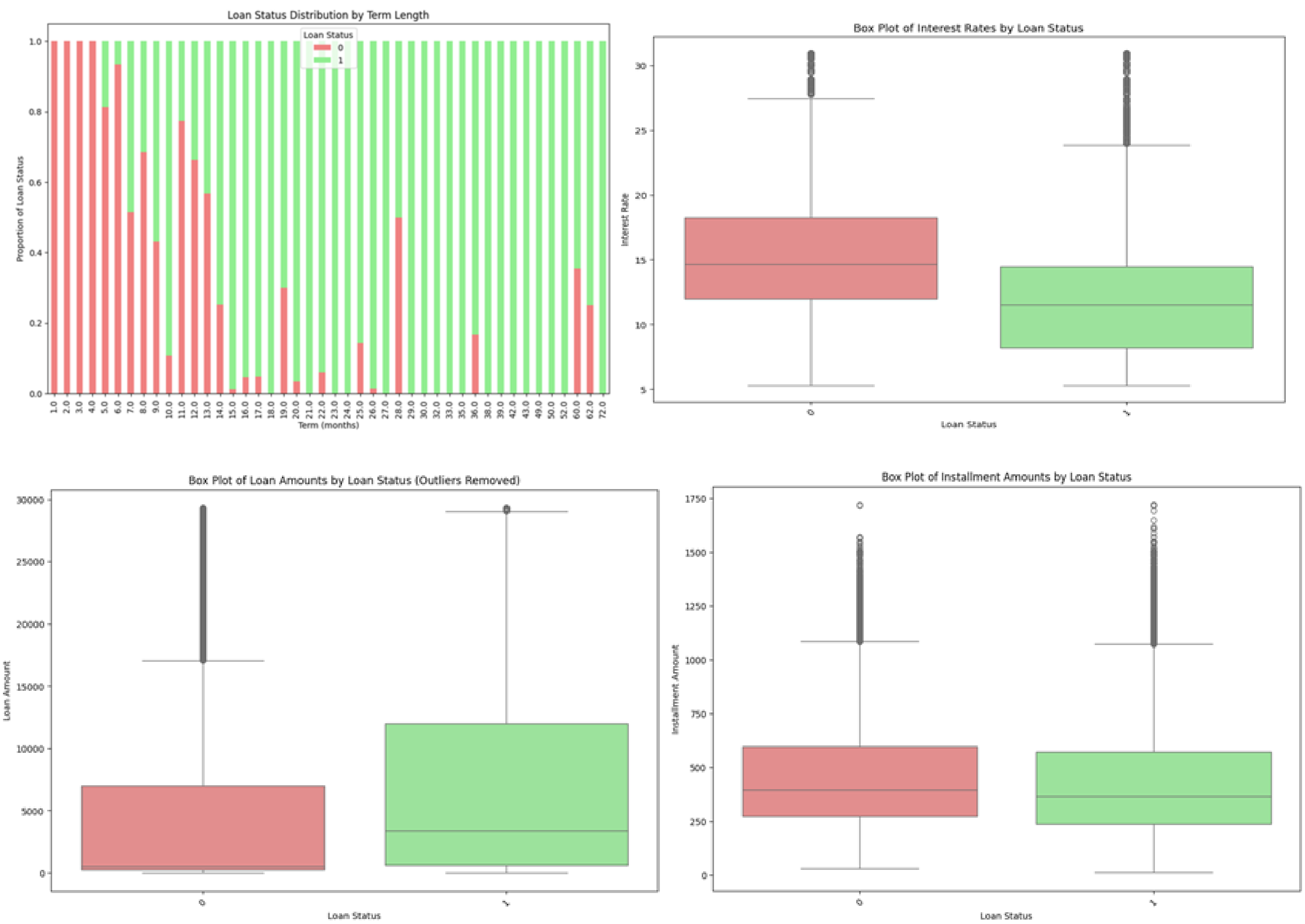

Figure 5: Loan characteristics distribution

Loan characteristics such as term, loan amount, interest rate, and instalment amount provide additional insights. Short-term loans (1-12 months) exhibit higher default rates, whereas longer-term loans are predominantly non-defaulted. Higher loan amounts are generally associated with non-defaulted loans, while interest rates tend to be higher for defaulted loans. Installments are also slightly higher in defaulted loans, indicating a broader range, but with minor differences between defaulted and non-defaulted loans (Figure 5). Using t-tests for numerical variables and chi-square tests for categorical variables, each feature yielded a highly significant result (p-value = 0.0), indicating that the observed relationships are not due to random chance. The results are as follows: Loan amount, interest rate, instalment amount, and annual income all had substantial t-statistics (ranging from approximately 99 to 152), reinforcing their importance in the model. High t-statistics and zero p-values suggest that variations in these numerical features are significantly associated with credit risk outcomes. Categorical variables, including gender, address state, home ownership, and loan term, were analysed using chi-square tests. All tests returned very high chi-square values (ranging from approximately 18,603 to 445,991) with p-values of zero, indicating significant associations between these features and credit risk. These results underscore the influence of these categorical factors on credit behaviour.

Table 2: Statistical validation of features

Feature	Test	Statistic	p-value	Significance
Interest Rate	t-test	99.47	0.0	Significant
Loan Amount	t-test	139.30	0.0	Significant
Installment	t-test	144.60	0.0	Significant
Annual Income	t-test	152.44	0.0	Significant
Gender	Chi-square	105471.23	0.0	Significant
Address State	Chi-square	445991.39	0.0	Significant
Home Ownership	Chi-square	48394.41	0.0	Significant
Term	Chi-square	18603.04	0.0	Significant

This statistical validation highlights the robustness of the features detailed in Table 2 above.

4.2. Model Development and Evaluation

Following the exploratory data analysis, five machine learning models were trained to predict credit risk under varying data conditions. Evaluation metrics included accuracy, precision, recall, and F1 score.

4.2.1. Model Performance

Table 3 presents a Model evaluation comparing five models—logistic regression, decision trees, random forests, gradient boosting, and extreme gradient boosting—across accuracy, precision, recall, and F1-score (for both classes and averages). Extreme gradient boosting ranks 1st (Avg. F1 ≈ 0.96), followed by random forest (2nd), decision trees (3rd), gradient boosting (4th), and logistic regression (5th).

Table 3: Model evaluation

Model Number	Model Name	Accuracy	Precision (Class 0)	Precision (Class 1)	Avg. Precision	Recall (Class 0)	Recall (Class 1)	Avg. Recall	F1 Score (Class 0)	F1 Score (Class 1)	Avg.F1 Score	Rank
1	Logistic Regression	0.8	0.63	0.93	0.78	0.86	0.78	0.82	0.73	0.85	0.79	5
2	Decision Trees	0.96	0.92	0.98	0.95	0.96	0.96	0.96	0.94	0.97	0.96	3
3	Random Forest	0.97	0.93	0.98	0.95	0.96	0.97	0.96	0.94	0.97	0.96	2
4	Gradient Boosting	0.92	0.84	0.97	0.9	0.93	0.92	0.93	0.88	0.94	0.91	4
5	Extreme Gradient Boosting	0.97	0.92	0.99	0.95	0.97	0.97	0.97	0.95	0.98	0.96	1

From Table 3 above, it can be observed that both Random Forest and Extreme Gradient Boosting attained the greatest accuracy of 97%. However, Extreme Gradient Boosting marginally outperformed in terms of precision, recall, and F1 Score. This places Extreme Gradient Boosting as the best-performing model among the five models evaluated. Although Decision Trees and Gradient Boosting did not outperform other ensemble approaches, they have obtained quite high accuracy. Lastly, Logistic Regression achieved lower overall accuracy; however, it still demonstrated strong precision for class 1 predictions. Since

Extreme Gradient Boosting and Random Forest were observed to perform consistently better than the others, the following documentation will focus only on these two models to provide a concise presentation of the results.

4.2.2. Extreme Gradient Boosting

`

Figure 6 shows how well the XGBoost model performs using a confusion matrix, an ROC curve (AUC = 0.99), and a plot comparing predicted and actual labels. It shows that the categorisation works well because there are many real positives and negatives, and the forecasts and actual values are very close.

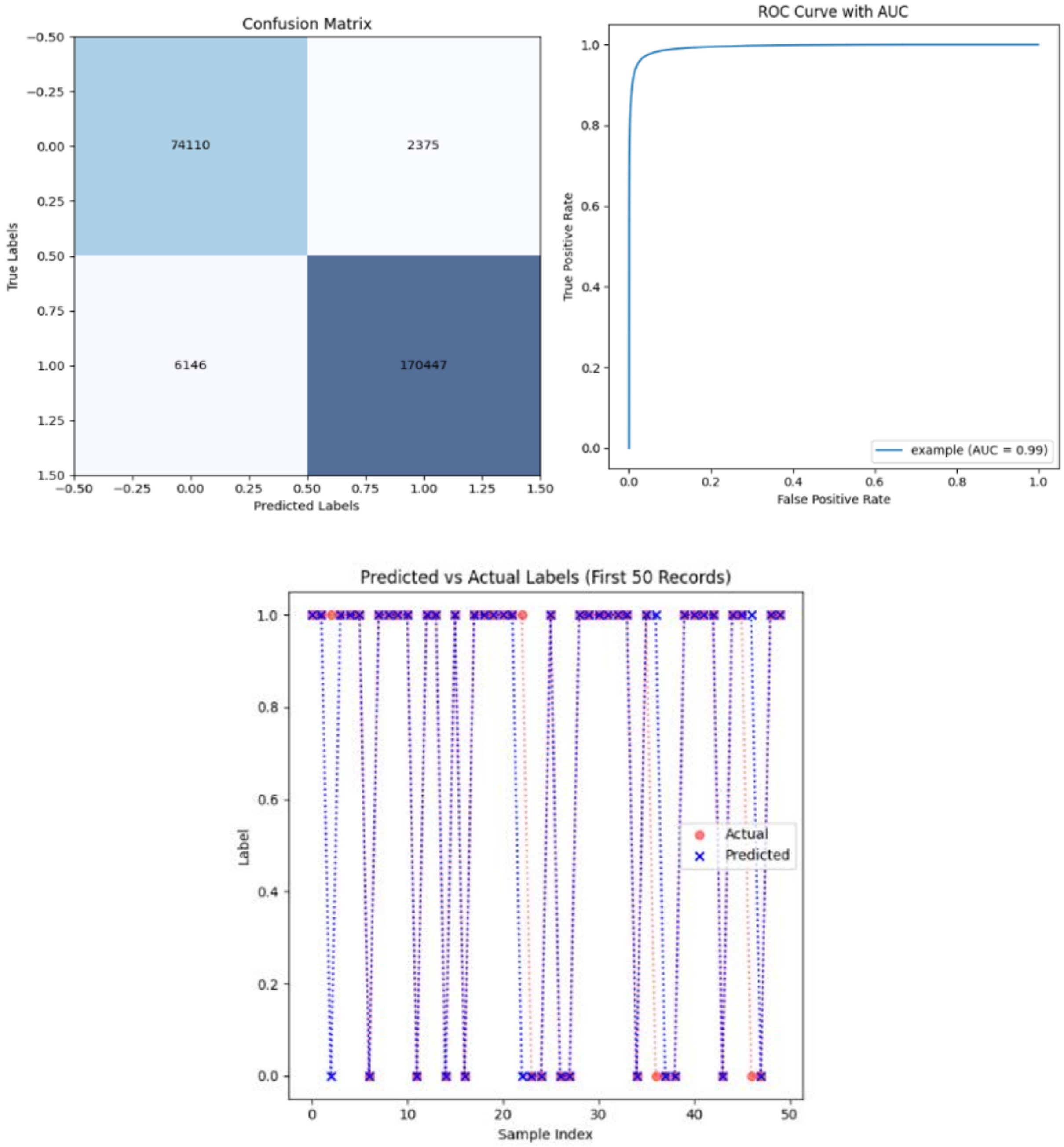

Figure 6: XGBoost model performance visualisation

Referring to Figure 6 above, the confusion matrix on the left shows the model's predictive performance for correct (true positive and true negative) and incorrect (false positive and false negative) predictions. XGBoost predicts 74,110 true negatives (*correctly predicted non-defaults*) and 170, 447 true positives (*correctly predicted defaults*). However, there are also 6146 false negatives and 2375 false positives showing incorrect predictions by the model. Additionally, the ROC curve in the center visualizes the effectiveness of the true positive rate and the false positive rate. The high AUC of 0.99 indicates that XGBoost effectively distinguishes between default and non-default loans. Furthermore, the image also shows the predicted and actual values for the first 50 samples. Figure 6, on the extreme left, shows how effective XGBoost is at classifying profiles with very few inaccuracies.

4.2.3. Random Forest

Similarly, Figure 7 shows the performance of Random Forest. The confusion matrix (*extreme left*) shows that the model correctly predicted 73,624 non-defaults and 170,447 defaults. The model incorrectly classified 8521 loans.

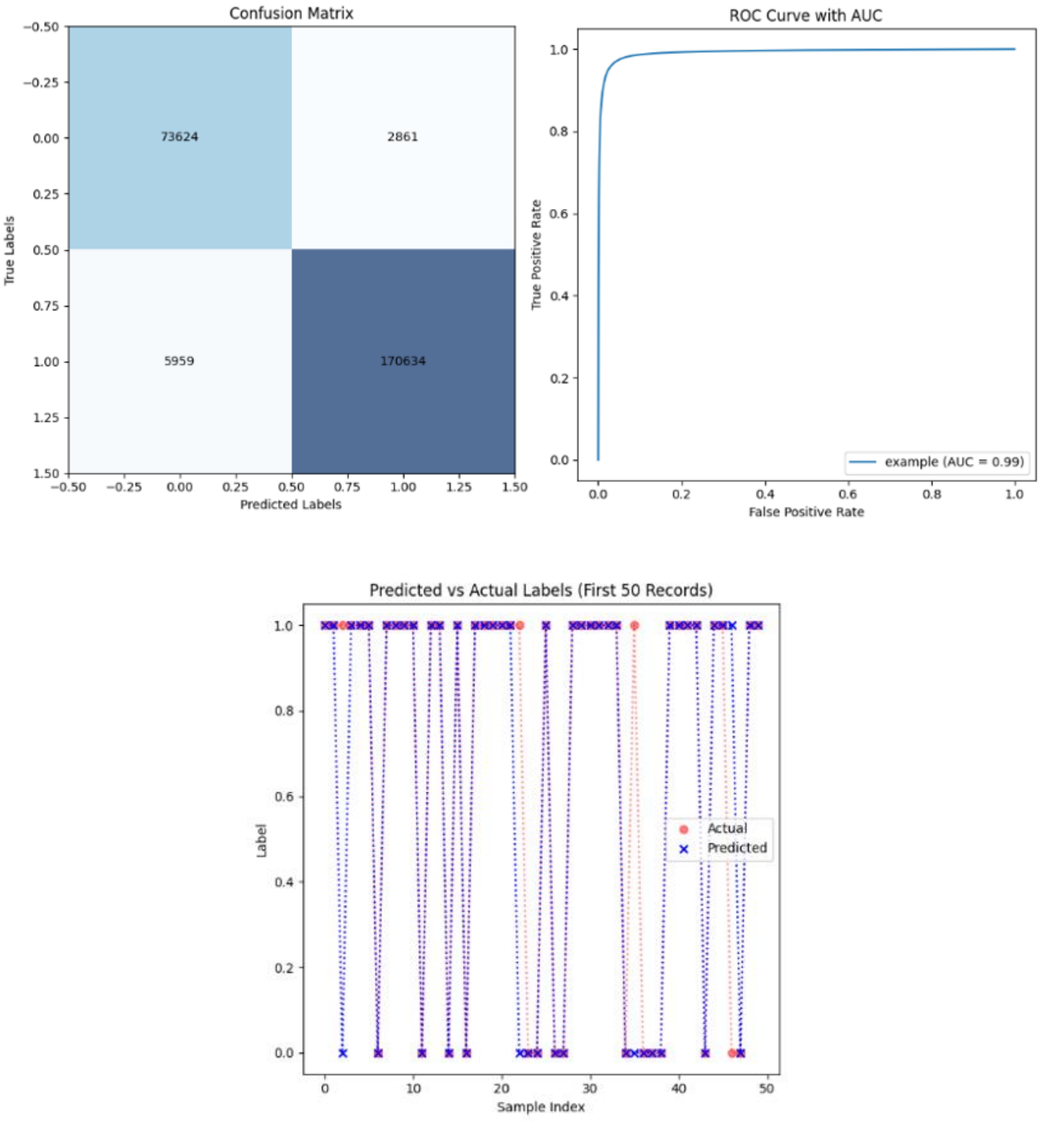

Figure 7: Random forest model performance visualisation

The central image demonstrates the models' ability to distinguish between defaulted and non-defaulted loans. Similar to XGBoost, this model also has an AUC of 0.99, which indicates very high effectiveness [47]. The chart on the right compares the predicted loan status with the actual loan status for the first 50 records. The few discrepancies observed in Figure 7 reflect the models' minor inaccuracies. Overall, both XGBoost and Random Forests show excellent predictive performance, accurately classifying data points. However, XGBoost shows slightly better performance in terms of precision, recall and F1 Score, which makes it a more effective model in this study. This paper presented the results of a comparative analysis of ML models and key factors in credit risk assessment. The results showed that ensemble methods such as Random Forest and Extreme Gradient Boosting performed better than traditional models. However, XGBoost showed more consistent performance across metrics, thus having slightly better performance than Random Forest. Key factors, including loan characteristics and borrower demographics, were also analysed in this research. These features include gender, employment sector, home ownership, annual income, term period, installment amount, interest rate, and loan amount. These findings show that the analysed features significantly influence loan status. These results align with the current literature, which will be discussed in the following paper. Furthermore, implications of these findings and any recommendations for practical implementation will also be discussed in the following section.

5. Discussion

These results support previous research by highlighting the effectiveness of ensemble approaches in credit risk assessment. These models work exceptionally well as they can manage intricate, non-linear relations in the data [23]. Thus, making them well-suited for tasks where the decision boundaries are ambiguous [34]. With these benefits, Ensemble methods (*such as XGBoost and Random Forest*) outperformed other traditional models in this research. These factors may have contributed to XGBoost's performance in this dataset. Consequently, having improved accuracy, recall, precision and F1 score compared to other models. This research has also identified that loan characteristics and borrower demographics influence default probability. Similar to the findings by Ampountolas et al. [4], this research confirmed that the interest rate is a critical predictor of credit risk. As argued by Wondirad [48] high interest rates lead to more defaults, thereby impacting the organisation's sustainability. In addition to the interest rate, this research has found that the borrower's instalment payments significantly affect the default probability. This research aligns with Durango-Gutiérrez et al. [14] in finding that female borrowers negatively affect credit risk. Researchers like Zainuddin and Yasin [51] argue that female borrowers positively influence an organisation's profit.

This difference in findings could potentially be due to the longitudinal nature of their research. The dataset's narrower time period might fail to capture this association. However, the higher default rate among female borrowers may also suggest that women face more financial challenges, such as lower average income or greater family care responsibilities [9]. Medina-Olivares et al. [32] concluded that loan size and location influence loan defaults. Although this study observed loan amount and geographical location to be critical factors, the direction of association contradicts their conclusions. This study observes that lower loan amounts are more subject to default. However, this might be due to the data distribution in this research. Furthermore, the variation in default rates across regions might also have been due to cultural or ethnic standards associated with loans. The conclusions also confirm the theory established by Medina-Olivares et al. [32] regarding the repayment term structure. Both analyses revealed that a longer repayment structure leads to lower default rates. In addition to these findings, this research identified an association with home ownership. Borrowers with no specific home status demonstrated lower default rates. Furthermore, the analysis has identified that the borrower's annual income affects repayment behaviour.

5.1. Theoretical and Practical Implications

This research provided evidence supporting the use of ensemble methods for risk assessment modeling. The research also identified key features for credit risk prediction, extending the current body of literature on risk predictors. Although the findings contradict some previous conclusions, they give the possibility for further research. Moreover, the findings also provide practical applications for financial institutions. These include:

- The organisations can consider implementing ensemble methods, especially Random Forest and XGBoost. These models have shown excellent performance.
- Lenders might develop tailored products or support programs to educate borrowers. Particularly, borrowers from vulnerable populations, such as females.
- Lenders could also develop policies that support region-specific challenges. They can also develop mitigation strategies based on economic conditions or behaviour.
- Institutions could benefit from providing extended repayment options or longer-term periods to reduce the risk of default.

5.2. Limitations

Although the analysis provided valuable insights, as discussed in the previous section, it also has limitations. These limitations might affect the impact of the findings. They are:

- Skewness in the data sample might have introduced bias in the model predictions or analysis.
- Although the data covered diverse countries, it had high density in the United States. This limits the applicability of the findings to the other countries.

- The sample focused only on a specific time period, which may not have analyzed long-term trends in borrower behavior.

Therefore, future research in this field could include a wider range of countries or regions. They can also conduct longitudinal research to examine changes or trends in borrower behaviour. Given the data skew, future researchers could focus on developing techniques that prioritise fairness. The field of credit risk might also benefit from studying the macroeconomic factors that influence default rates.

5.3. Unexpected Results

The most unexpected finding in this research was the high default rate with smaller loan amounts. This result contradicts the common belief that smaller loans are less risky. However, this result might indicate that borrowers who take smaller loans are more financially unstable. Another possible explanation is the influence of subjective factors, such as political stability or economic policies. However, these factors are out of scope for this research.

6. Conclusion

The study initially aimed to evaluate five ML models for predicting loan defaults under diverse data conditions. The models assessed in this study are logistic regression, decision trees, random forest, gradient boosting and extreme gradient boosting. The models were trained and tested using a combined dataset from two different lending platforms. The metrics used to assess their performance are accuracy, precision, recall, F1 score, confusion matrix and ROC AUC. To demonstrate the model's efficiency, the research also visualised the predicted and actual values for the first 50 data samples. The analysis also aimed to identify the factors that influence default probability. The features included in this analysis were the readily available information from a borrower or lender. The selected features include details about the borrower and the loan. These features include gender, annual income, home ownership status, location, loan amount, interest rate, installment, and term period. These objectives were formulated to understand the research questions, which are as follows:

- Which ML model provides the most accurate predictions of credit risk, even with data limitations?
- What loan characteristics and social factors affect the performance of a loan and borrowers' behaviour?

The results showed that ensemble models, particularly XGBoost and Random Forest, performed well. These models demonstrated consistent performance across all the metrics. However, Extreme Gradient Boosting performed slightly better in some of them. These factors may have contributed to this superior performance. Thus, making them well-suited for tasks of predicting risky credit profiles. Through data analysis, female borrowers, individuals with low annual income, and tenants were found to be at higher risk. Furthermore, profiles based in the United States and the Philippines showed high default risk. On the other hand, small loan amounts, high interest rates, short loan periods and high installments were associated with high default rates. These conclusions, although mostly in line with the current literature, could be further supported by additional validations. These findings confirm the existing theories in credit risk assessment. They also provide insights for organizations seeking to refine their credit risk procedures. Some of these practical insights may include tailored products or support systems for underrepresented groups, region- or country-specific strategies, and so on. Even with these results, the study had its own limitations. One limitation is a highly skewed distribution, which can limit the generalizability of the findings. Furthermore, the sample focused on a short time frame, which did not allow for the study of trends in borrower behaviour. Given these limitations, future researchers could extend this field by studying qualitative and macroeconomic factors and their influence on credit risk. The studies can also expand the dataset with a broader population and conduct more diverse analyses. The researchers could also study strategies that incorporate fairness in ML models.

Acknowledgment: The authors would like to acknowledge Northumbria University for its support and resources, which contributed to the completion of this study.

Data Availability Statement: Data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to applicable permissions.

Funding Statement: This study was conducted without any external financial support.

Conflicts of Interest Statement: The authors declare that they have no conflicts of interest. All sources used in this original work have been properly acknowledged.

Ethics and Consent Statement: This study was conducted in accordance with established ethical guidelines. Informed consent was obtained from all participants before their participation.

References

1. R. Alasbahi and X. Zheng, "An online transfer learning framework with extreme learning machine for automated credit scoring," *IEEE Access*, vol. 10, no. 5, pp. 46697–46716, 2022.
2. A. Alonso-Robisco and J. M. Carbó, "Can machine learning models save capital for banks? Evidence from a Spanish credit portfolio," *International Review of Financial Analysis*, vol. 84, no. 11, p. 102372, 2022.
3. L. Alzubaidi, J. Bai, A. Al-Sabaawi, J. Santamaria, A. S. Albahri, B. S. N. Al-Dabbagh, M. A. Fadhel, M. Manoufali, J. Zhang, A. H. Al-Timemy, Y. Duan, A. Abdullah, L. Farhan, Y. Lu, A. Gupta, F. Albu, A. Abbosh, and Y. Gu, "A survey on deep learning tools dealing with data scarcity: Definitions, challenges, solutions, tips, and applications," *Journal of Big Data*, vol. 10, no. 1, p. 46, 2023.
4. A. Ampountolas, T. N. Nde, P. Date, and C. Constantinescu, "A machine learning approach for micro-credit scoring," *Risks*, vol. 9, no. 3, p. 50, 2021.
5. R. A. Berk, A. K. Kuchibhotla, and E. T. Tchetgen, "Fair risk algorithms," *Annual Review of Statistics and Its Application*, vol. 10, no. 1, pp. 165–187, 2023.
6. T. Bono, K. Croxson, and A. Giles, "Algorithmic fairness in credit scoring," *Oxford Review of Economic Policy*, vol. 37, no. 3, pp. 585–617, 2021.
7. S. Caton, S. Malisetty, and C. Haas, "Impact of imputation strategies on fairness in machine learning," *Journal of Artificial Intelligence Research*, vol. 74, no. 6, pp. 1011–1035, 2022.
8. S. Chen, "A novel ensemble machine learning model for credit risk prediction," *Advances in Economics, Management and Political Sciences*, vol. 46, no. 12, pp. 277–281, 2023.
9. H. I. Condori-Alejo, M. R. Aceituno-Rojo, and G. S. Alzamora, "Rural micro credit assessment using machine learning in a Peruvian microfinance institution," *Procedia Computer Science*, vol. 187, no. 1, pp. 408–413, 2021.
10. S. Cornée, "The relevance of soft information for predicting small business credit default: Evidence from a social bank," *Journal of Small Business Management*, vol. 57, no. 3, pp. 699–719, 2019.
11. A. Danandeh Mehr and N. Fathollahzadeh Attar, "A gradient boosting tree approach for SPEI classification and prediction in Turkey," *Hydrological Sciences Journal*, vol. 66, no. 11, pp. 1653–1663, 2021.
12. A. De Hond, W. Raven, L. Schinkelshoek, M. Gaakeer, E. Ter Avest, O. Sir, H. Lameijer, R. A. Hessels, R. Reijnen, E. De Jonge, E. Steyerberg, C. H. Nickel, and B. De Groot, "Machine learning for developing a prediction model of hospital admission of emergency department patients: Hype or hope?," *International Journal of Medical Informatics*, vol. 152, no. 8, pp. 1–8, 2021.
13. E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi, "Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects," *European Journal of Operational Research*, vol. 297, no. 3, pp. 1178–1192, 2022.
14. M. P. Durango-Gutiérrez, J. Lara-Rubio, and A. Navarro-Galera, "Analysis of default risk in microfinance institutions under the Basel III framework," *International Journal of Finance & Economics*, vol. 28, no. 2, pp. 1261–1278, 2023.
15. A. A. A. Fadikpe, R. Danquah, M. Aidoo, D. A. Chomen, R. Yankey, and X. Dongmei, "Linkages between social and financial performance: Evidence from Sub-Saharan Africa microfinance institutions," *PLOS ONE*, vol. 17, no. 3, pp. 1–23, 2022.
16. J. Fang and J. Zhu, "The impact of open banking on traditional lending in the BRICS," *Finance Research Letters*, vol. 58, no. 12, p. 104300, 2023.
17. R. Guo, D. Fu, and G. Sollazzo, "An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree," *International Journal of Pavement Engineering*, vol. 23, no. 10, pp. 3633–3646, 2022.
18. A. A. H. A. Khatir and M. Bee, "Machine learning models and data-balancing techniques for credit scoring: What is the best combination?," *Risks*, vol. 10, no. 9, p. 169, 2022.
19. G. M. Jakka, A. Panigrahi, A. Pati, M. N. Das, and J. Tripathy, "A novel credit scoring system in financial institutions using artificial intelligence technology," *Journal of Autonomous Intelligence*, vol. 6, no. 2, pp. 1–15, 2023.
20. R. Jarrow and P. Protter, "Fair microfinance loan rates," *International Review of Finance*, vol. 19, no. 4, pp. 909–918, 2019.
21. C. Johnen, M. Parlasca, and O. Mußhoff, "Promises and pitfalls of digital credit: Empirical evidence from Kenya," *PLOS ONE*, vol. 16, no. 7, pp. 1–14, 2021.
22. M. Jumaa, M. Saqib, and A. Attar, "Improving credit risk assessment through deep learning-based consumer loan default prediction model," *International Journal of Finance & Banking Studies*, vol. 12, no. 1, pp. 85–92, 2023.
23. V. Kanaparthi, "Credit risk prediction using ensemble machine learning algorithms," in *2023 International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, 2023.

24. G. Klein, Z. Shtudiner, and M. Zwilling, “Why do peer-to-peer (P2P) lending platforms fail? The gap between P2P lenders’ preferences and the platforms’ intentions,” *Electronic Commerce Research*, vol. 23, no. 6, pp. 709–738, 2023.
25. J. M. Klusowski and P. M. Tian, “Large scale prediction with decision trees,” *Journal of the American Statistical Association*, vol. 119, no. 545, pp. 525–537, 2024.
26. N. Kozodoi, J. Jacob, and S. Lessmann, “Fairness in credit scoring: Assessment, implementation and profit implications,” *European Journal of Operational Research*, vol. 297, no. 3, pp. 1083–1094, 2022.
27. B. Li, B. Xiao, and Y. Yang, “Strengthen credit scoring system of small and micro businesses with soft information: Analysis and comparison based on neural network models,” *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, vol. 40, no. 3, pp. 4257–4274, 2021.
28. D. Li, “Evaluating various machine learning techniques in credit risk area,” *BCP Business & Management*, vol. 38, no. 3, pp. 2836–2844, 2023.
29. H. Lin, Y. Xue, K. Chen, S. Zhong, and L. Chen, “Acute coronary syndrome risk prediction based on gradient boosted tree feature selection and recursive feature elimination: A dataset-specific modeling study,” *PLOS ONE*, vol. 17, no. 11, pp. 1–24, 2022.
30. M. Lusinga, T. Mokoena, A. Modupe, and V. Mariate, “Investigating statistical and machine learning techniques to improve the credit approval process in developing countries,” in *Proc. 2021 IEEE AFRICON*, Arusha, Tanzania, 2021.
31. N. T. Luu and P. D. Hung, “Loan default prediction using artificial intelligence for the borrow–lend collaboration,” in Cooperative Design, Visualization, and Engineering, Y. Luo, Ed. *Springer International Publishing*, Cham, Switzerland, 2021.
32. V. Medina-Olivares, R. Calabrese, Y. Dong, and B. Shi, “Spatial dependence in microfinance credit default,” *International Journal of Forecasting*, vol. 38, no. 3, pp. 1071–1085, 2022.
33. A. K. Mishra and V. Bhardwaj, “Financial access and household’s borrowing: Policy perspectives of an emerging economy,” *Journal of Policy Modeling*, vol. 44, no. 5, pp. 981–999, 2022.
34. J. Nalić, Z. Mašetić, and I. Djedović, “Building ensemble models for enhanced credit scoring: A case study from a Bosnian microfinancial institution,” in *Proc. 2024 47th MIPRO ICT and Electronics Convention (MIPRO)*, Opatija, Croatia, 2024.
35. J. Nalić, Z. Mašetić, and I. Djedović, “Enhancing credit scoring models: Unveiling the impact of data pre-processing and feature selection techniques: Case study based on microfinancial institution in Bosnia and Herzegovina,” in *Proc. 2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH)*, East Sarajevo, Bosnia and Herzegovina, 2024.
36. J. P. Noriega, L. A. Rivera, and J. A. Herrera, “Machine learning for credit risk prediction: A systematic literature review,” *Data*, vol. 8, no. 11, pp. 1–17, 2023.
37. P. Paraíso, S. Ruiz, P. Gomes, L. Rodrigues, and J. Gama, “Using network features for credit scoring in microfinance,” *International Journal of Data Science and Analytics*, vol. 12, no. 2, pp. 121–134, 2021.
38. K. D. Pradnyana, “Loan default prediction in microfinance group lending with machine learning,” Master’s thesis, *Institut Teknologi Bandung*, Bandung, Indonesia, 2022.
39. P. K. Roy and K. Shaw, “A multicriteria credit scoring model for SMEs using hybrid BWM and TOPSIS,” *Financial Innovation*, vol. 7, no. 10, pp. 1–27, 2021.
40. M. N. K. Saunders, “Understanding research philosophies and approaches,” in *Research Methods for Business Students*, vol. 4, no. 1, pp. 121–161, 2009.
41. S. Shi, R. Tse, W. Luo, S. D’Addona, and G. Pau, “Machine learning-driven credit risk: A systemic review,” *Neural Computing and Applications*, vol. 34, no. 7, pp. 14327–14339, 2022.
42. R. Shukla, R. Sawant, and R. Pawar, “A comparative study of deep learning and machine learning techniques in credit score classification,” *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 11, no. 7, pp. 10004–10010, 2023.
43. M. Strobel and R. Shokri, “Data privacy and trustworthy machine learning,” *IEEE Security & Privacy*, vol. 20, no. 5, pp. 44–49, 2022.
44. N. Suhadolnik, J. Ueyama, and S. Da Silva, “Machine learning for enhanced credit risk assessment: An empirical approach,” *Journal of Risk and Financial Management*, vol. 16, no. 12, pp. 1–21, 2023.
45. K. Sytnychenko, “What to know before taking out a payday loan to improve your credit score,” *International Science Journal of Management, Economics & Finance*, vol. 2, no. 2, pp. 10–16, 2023.
46. P. Tumuluru, L. R. Burra, M. Loukya, S. Bhavana, H. M. H. CSaiBaba, and N. Sunanda, “Comparative analysis of customer loan approval prediction using machine learning algorithms,” in *Proc. 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, Tamil Nadu, India, 2022.
47. C. Wang and Z. Xiao, “A deep learning approach for credit scoring using feature embedded transformer,” *Applied Sciences*, vol. 12, no. 21, pp. 1–14, 2022.
48. H. A. Wondirad, “Interest rates in microfinance: What is a fair interest rate when we lend to the poor?,” *Quality & Quantity*, vol. 56, no. 2, pp. 4537–4548, 2022.

49. B. Wu, Z. Liu, Q. Gu, and F. S. Tsai, “Underdog mentality, identity discrimination and access to peer-to-peer lending market: Exploring effects of digital authentication,” *Journal of International Financial Markets, Institutions and Money*, vol. 83, no. 3, p. 101714, 2023.
50. F. M. Zahid, S. Faisal, and C. Heumann, “Multiple imputation with compatibility for high-dimensional data,” *PLOS ONE*, vol. 16, no. 7, pp. 1–15, 2021.
51. M. Zainuddin and I. M. Yasin, “Are women better borrowers in microfinance? A global analysis,” *The Empirical Economics Letters*, vol. 19, no. 7, pp. 651–660, 2020.