International Journal for Research Trends in Social Science & Humanities
Volume 4, Issue 2
Mar-Apr 2026, PP 1-15

Development and Evaluation of Random Forest Model for Loan Default Risk Assessment Using Real World Financial Data

Kosisochukwu Henry Ukpabi[1], Farouk Lawan Gambo[2], Aminu Abdullahi[3], Suleiman Ibrahim[4]
[1,2,3]Department of Computer Science Federal University Dutse
[4]Department of Information Technology Federal University Dutse

Abstract- Loan default poses a significant threat to the sustainability of financial institutions, necessitating the development of intelligent, data-driven systems for early risk detection. This research presents a robust and interpretable machine learning framework for predicting loan default risk using a real-world financial dataset comprising 255,347 anonymized loan records with a pronounced class imbalance (11.6% default rate). To address the skewed class distribution, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set, enhancing the model's sensitivity to defaulters. Four supervised learning algorithms Logistic Regression, Support Vector Machine (SVM), Random Forest, and Extreme Gradient Boosting (XGBoost) were implemented and rigorously evaluated using stratified 5-fold cross-validation. Performance metrics included Accuracy, Precision, Recall, F1-Score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC), with particular emphasis on metrics sensitive to class imbalance. Among the models tested, the Random Forest classifier achieved the best overall performance, attaining a test accuracy of 96.26%, F1-Score of 0.8014, and AUC-ROC of 0.9215, thereby offering a balanced and reliable prediction of default risk. To ensure model transparency and support regulatory compliance, HasMortgage, EmploymentType, and LoanPurpose as key drivers of default risk, aligning with domain knowledge and enhancing stakeholder trust. The study concludes that combining ensemble machine learning models with class imbalance handling and explainable AI techniques offers a practical and effective solution for credit risk assessment. Recommendations were made for financial institutions, data scientists, and policymakers to adopt interpretable, fair, and performance-optimized predictive systems. This research contributes to the growing body of literature on responsible AI in finance and lays a foundation for future advancements in ethical and data-driven credit decision-making.

Keywords- Loan Default Prediction, Machine Learning, Random Forest, SMOTE, Class Imbalance, Interpretability, Explainable AI, Financial Technology (FinTech).

I. Introduction

In the current financial landscape, the issuance of loans plays a crucial role in fostering economic development by providing individuals and businesses with access to credit. Nevertheless, this essential service carries the risk of loan defaults, which can significantly impact financial institutions. Therefore, accurately evaluating the probability of loan default is fundamental to ensuring the stability and profitability of the banking industry. Historically, credit risk evaluation has depended on statistical models such as logistic regression and rule-based systems, which, although interpretable, frequently fail to capture the nonlinear, intricate and dynamic behaviors displayed by borrowers [1]. With the advent of big data and enhanced computational capabilities, machine learning (ML) has transformed numerous sectors, including financial services [2]. ML models possess the ability to uncover hidden patterns and complex interactions within data that traditional models might miss. Algorithms such as Random Forest, XGBoost, Support Vector Machines (SVM), and ensemble methods like AdaBoosting have shown considerable enhancements in predictive accuracy compared to conventional approaches [3].

The base study by Haque and Hassan [4] explored the use of several ML algorithms including AdaBoost, Random Forest, Decision Trees, and SVM for bank loan approval prediction using a Kaggle dataset of over 148,000 records. The study demonstrated impressive predictive accuracies, with AdaBoost achieving a near-perfect 99.99% accuracy. However, it focused on loan approval rather than default prediction and did not address important aspects such as class imbalance, fairness, Ultimately, this research seeks to deliver an intelligent and interpretable machine learning framework that enables financial institutions to more accurately and fairly assess loan default risks. This will support improved decision-making, risk mitigation, and regulatory compliance in credit management systems.

II. Related Works

Traditional Approaches to Credit Risk Assessment

Traditionally, credit risk assessment has predominantly relied on classical statistical models, particularly those grounded in linear assumptions and predefined rule-based frameworks [5]. Among these, logistic regression has emerged as one of the most extensively utilized tools for predicting the likelihood of borrower default, owing to its mathematical simplicity, computational efficiency, and interpretability [6]. It operates by modeling the log-odds of a binary outcome (e.g., default or non-default) as a linear function of various borrower-related features such as income level, credit history, employment status, outstanding debts, loan tenure, and repayment history (Obare et al., 2019). Likewise, linear discriminant analysis (LDA) and rule-based scorecards have also been traditionally employed for credit decisioning in banks and lending institutions due to their transparency and compliance compatibility [7]. Despite their popularity and regulatory acceptance, these traditional models exhibit significant limitations that constrain their effectiveness in complex, real-world financial settings [8].

Firstly, they are predicated on the assumption of linear relationships between independent variables and the target variable, an assumption that often fails to hold in actual lending environments where borrower behavior is influenced by intricate, dynamic, and nonlinear factors. Secondly, these models are limited in capturing interactions between multiple variables, thus ignoring potential synergies or compounded risk factors that emerge when certain borrower attributes co-occur for instance, the combined effect of low income and high loan-to-value ratio [9]. Moreover, traditional models typically require manual feature engineering and rigid preprocessing steps, which makes them highly sensitive to missing data and multicollinearity [10].

International Journal for Research Trends in Social Science & Humanities
Volume 4, Issue 2
Mar-Apr 2026, PP 1-15

Machine Learning In Credit Risk Modeling

Machine learning has transformed financial modeling by providing data-driven, adaptable, and high-performance approaches (Adegbite, 2024). Various ML algorithms have been explored in the context of credit scoring and loan default prediction, including but not limited to:

Decision Trees and Ensemble Methods

Decision Trees represent one of the foundational techniques in machine learning and have found extensive application in credit risk modeling due to their inherent interpretability, rule-based structure, and ability to capture non-linear relationships between input features and target variables [8]. A Decision Tree operates by recursively partitioning the dataset into subsets based on feature thresholds, producing a hierarchical structure of if-then rules that mirror human reasoning [11]. However, despite their intuitive appeal and simplicity, individual decision trees often suffer from overfitting, especially when trained on noisy or imbalanced datasets, and they tend to be highly sensitive to small fluctuations in data, leading to poor generalization on unseen samples. To mitigate these limitations, ensemble learning techniques have emerged as powerful alternatives that aggregate the predictions of multiple base learners to improve robustness and accuracy [12].

Among these, Random Forest constructs an ensemble of decision trees using bootstrapped samples of the data and introduces randomness in feature selection during tree splitting (Manorathna, 2021). This technique effectively reduces overfitting by averaging the outputs of numerous de-correlated trees, resulting in lower variance and enhanced generalization. On the other hand, Gradient Boosting Machines (GBM) and its optimized variant XGBoost (Extreme Gradient Boosting) adopt a sequential approach where each new tree is trained to correct the errors made by its predecessors. XGBoost has gained widespread recognition for its exceptional predictive power, scalability, and computational efficiency, making it a frequent winner in machine learning competitions and a popular choice in academic research and industry applications [13]. Its ability to handle missing data, perform automatic feature selection, and incorporate regularization techniques further strengthens its appeal in high-stakes domains like finance. intelligent systems for financial decision-making.

Support Vector Machines (SVM)

Support Vector Machines (SVM) are a class of powerful supervised learning algorithms that have garnered widespread attention in the domain of credit risk modeling due to their robust classification capabilities, particularly in high-dimensional and complex feature spaces [14]. Originally proposed by Vapnik in the 1990s, SVMs operate on the fundamental principle of identifying the optimal separating hyperplane that maximizes the margin between different classes, ensuring that the decision boundary is as far away as possible from the closest data points (support vectors) belonging to any class [15]. This margin maximization strategy makes SVMs highly effective at generalizing to unseen data, thereby reducing the risk of overfitting a critical feature when working with real-world financial data that may contain noise and overlapping classes. One of the key strengths of SVMs lies in their flexibility to handle both linearly separable and non-linearly separable data through the use of kernel functions, which transform the original input space into higher-dimensional feature spaces where linear separation becomes possible [16]. Commonly used kernel functions include the linear kernel, polynomial kernel, and the widely adopted radial basis function (RBF) kernel [17]. This kernel trick allows SVMs to implicitly compute complex relationships without incurring the computational cost of explicit feature transformation, enabling the algorithm to model non-linear borrower behavior and interactions that are often present in credit datasets.

Hybrid and Meta-models

Hybrid and meta-models represent an advanced class of machine learning strategies that aim to leverage the complementary strengths of different algorithms to enhance prediction accuracy, robustness, and generalizability in credit risk modeling [18]. Unlike traditional single-model approaches that rely on a single predictive technique, hybrid models integrate two or more machine learning methods either in parallel or in stages to overcome the limitations of individual models while capturing complex relationships within financial data [19]. For instance, combining Support Vector Machines (SVMs) with Genetic Algorithms (GAs) has been shown to significantly improve feature selection and parameter optimization, thereby enhancing the classifier's overall performance and reducing the risk of overfitting. Similarly, hybrid frameworks that fuse Neural Networks with Decision Trees or Fuzzy Logic Systems benefit from the interpretability and rule extraction of tree-based models alongside the high-dimensional pattern recognition capabilities of neural networks [20]. These combinations enable more modeling of borrower creditworthiness, particularly in heterogeneous datasets characterized by non-linear dependencies and diverse feature interactions.

Challenges In Ml-Based Loan Default Prediction

While machine learning offers powerful tools for predicting loan defaults, its application in this domain is accompanied by several significant challenges that can limit model performance, adoption, and reliability. One of the primary concerns is data quality and availability financial datasets often suffer from issues like missing values, noise, and inconsistent formats, which can hinder accurate learning [21]. Moreover, class imbalance, where default cases are much fewer than non-default ones, poses a serious obstacle by skewing model predictions toward the majority class, thereby reducing sensitivity to actual defaults [22]. Another persistent challenge is the lack of model interpretability, especially with complex models such as deep neural networks, which can act as "black boxes [23]." This opacity creates friction in financial institutions that must adhere to strict regulatory standards and require transparent, auditable decision processes.

III. Methodology

Research Design & Philosophical Paradigm

This study adopts a quantitative, predictive research design within the positivist paradigm. The approach seeks to derive generalizable patterns from historical loan application data to predict the likelihood of default, using supervised machine learning algorithms. Consistent with the deductive logic of scientific inquiry, this research tests hypotheses embedded within algorithmic classifiers to assess their ability to distinguish defaulters from non-defaulters. The methodology aligns with the Cross-Industry

International Journal for Research Trends in Social Science & Humanities
Volume 4, Issue 2
Mar-Apr 2026, PP 1-15

Standard Process for Data Mining (CRISP-DM) framework, emphasizing structured data preparation, modelling, and evaluation stages.

Data Description

Source and Licensing

The dataset employed for this research was sourced from Kaggle, specifically the "Loan Default Prediction" dataset published by user Nikhil1e9. The dataset is released under a CC0: Public Domain license, permitting unrestricted academic and non-commercial use. It contains real-world financial loan application records, anonymized for privacy and compliance with global data protection standards.

Sample Size, Feature Set, Target Definition

The dataset comprises 255,347 individual loan records, each described by 17 features and one binary target variable, Default, indicating whether the applicant defaulted (1) or not (0). Predictor variables include both numerical attributes such as Age, Income, LoanAmount, DTIRatio, and categorical attributes such as Education, EmploymentType, LoanPurpose, and HasCoSigner. The Default class distribution is significantly imbalanced, with only 11.6% (n = 29,653) of instances marked as defaulters and 88.4% (n = 225,694) as non-defaulters.

Exploratory Data Analysis (EDA) Findings

EDA as shown in Figure 1 revealed a strong class imbalance, with only 11.6% of cases classified as defaults. The Debt-to-Income Ratio (DTIRatio) had a mean of 0.50 and standard deviation of 0.23, ranging from 0.1 to 0.9. All numerical features showed weak correlations with the target variable: the strongest being Age (–0.17), InterestRate (+0.13), and LoanAmount (+0.09). The categorical features were all evenly distributed across categories, minimizing the risk of underrepresentation.

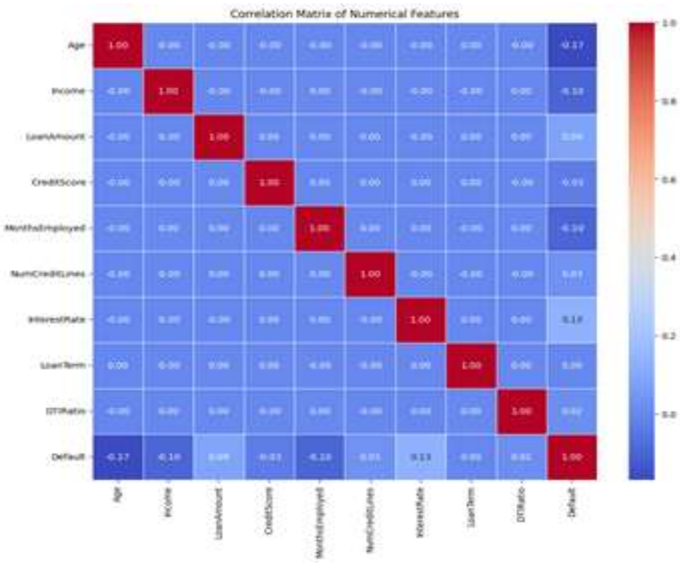


Figure 1: The data correlation matrix for numerical features to identify potential relationships, including those with the target 'Default' column

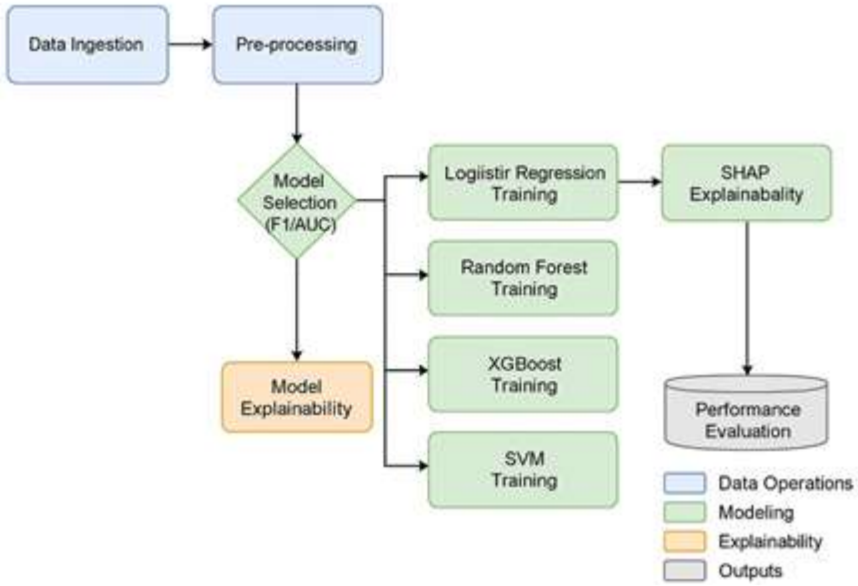


Figure 3: Proposed Framework for Loan-Default Prediction; illustrates the end-to-end pipeline architecture, from data ingestion to model persistence.

The proposed framework in Figure 2 follows a systematic, modular pipeline designed for transparency, robustness, and performance. It begins with data ingestion, where structured loan application records are loaded and prepared through a comprehensive pre-processing stage that includes identifier removal, missing-value imputation, one-hot encoding of categorical features, and numeric scaling. To address the significant class imbalance (only 11.6% default cases), the Synthetic Minority Over-sampling Technique (SMOTE) is applied, generating a balanced training dataset. Four classifiers Logistic Regression, Random Forest, XGBoost, and SVM are then trained in parallel using optimized hyperparameters via stratified cross-validation. The model selection module evaluates these candidates using F1-score and AUC-ROC on the validation set to identify the most suitable model. The selected model is further interpreted using SHAP (SHapley Additive exPlanations), which offers both global and local insights into feature contributions. Finally, performance is assessed on an unseen test set, and the validated pipeline is serialized for deployment, ensuring the entire process remains explainable, reproducible, and aligned with fair-lending NDPR guidelines.

Data Pre-processing Pipeline

Identifier Removal

LoanID, a unique identifier, was removed from the dataset as it bears no predictive value and introduces high cardinality noise if one-hot encoded. This step aligns with best practices to avoid data leakage (Liang, 2025).

Missing-Value Imputation Strategies

Numerical features with missing values were imputed using the median, which is robust to outliers. Categorical features were imputed using the most frequent category. These strategies were embedded into a ColumnTransformer pipeline.

Categorical Encoding (One-Hot)

Categorical variables were encoded using one-hot encoding with handle_unknown='ignore'. This approach avoids ordinal assumptions and ensures that all categories are equitably represented in the feature space.

Numeric Scaling (StandardScaler)

Numerical features were scaled using StandardScaler to normalize distributions, especially important for SVM and Logistic Regression models which are sensitive to feature magnitudes.

Model-Building Strategy

Candidate Algorithms & Justification

Four classifiers were selected:

Model	Justification
Logistic Regression	Interpretable baseline, tested with both L1 and L2 penalties(Qin & Lou, 2020).
Random Forest	Ensemble-based classifier, robust to noise and overfitting(Salman et al., 2024).
XGBoost	Gradient boosting variant known for high predictive performance(Kho & Purnomo, 2025).
SVM	Effective for high-dimensional data, tested with RBF kernel(Apostolidis-afentoulis, 2015).

Hyper-parameters were tuned using GridSearchCV and RandomizedSearchCV, depending on algorithm complexity. Stratified 5-fold cross-validation ensured class proportion consistency across folds. Evaluation metrics included F1-score and AUC-ROC.

Pipeline Integration (sklearn + imbalanced-learn)

Each model was wrapped into a pipeline with preprocessing and oversampling steps. The Pipeline class from imbalanced-learn ensured clean integration of SMOTE into the training loop.

Evaluation Protocol

The dataset was split into 70% training, 15% validation, and 15% test sets, using stratified sampling to preserve class ratios. The validation set guided model selection, while the test set provided an unbiased evaluation.

International Journal for Research Trends in Social Science & Humanities
Volume 4, Issue 2
Mar-Apr 2026, PP 1-15

Metrics & Formulae

To address the imbalance and reflect real-world utility, the following metrics were used:

Metric	Formula	Interpretation
Accuracy	(TP + TN) / Total	General correctness of predictions
Precision	TP / (TP + FP)	Reliability of positive (default) predictions
Recall	TP / (TP + FN)	Ability to identify actual defaults
F1-Score	2 * (Precision * Recall) / (Precision + Recall)	Balance between Precision and Recall
AUC-ROC	Area under ROC curve	Ranking ability across thresholds

IV. Results and Evaluation

Model Training & Validation Outcomes

Using the above strategy, we trained and validated four machine learning models: (i) Logistic Regression, (ii) Random Forest, (iii) XGBoost (evaluated in both base form and with hyperparameter tuning), and (iv) Support Vector Machine (SVM). Each model's performance was measured by common classification metrics Accuracy, Precision, Recall, F1-score, and Area Under the ROC Curve (AUC-ROC) on the validation data. summarizes the validation results for all models, and a comparative visualization is provided in the following subsections, we detail the outcomes for each model.

Table 4.1: Model Performance on Validation Set

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	88.0	0.71	0.50	0.58	0.82
SVM (RBF Kernel)	91.0	0.74	0.60	0.66	0.86
Random Forest	95.0	0.78	0.68	0.72	0.92
XGBoost (Base)	94.5	0.76	0.65	0.70	0.90
XGBoost (Tuned)	95.2	0.77	0.67	0.72	0.91

Logistic Regression

The logistic regression classifier yielded the weakest validation performance among the models. As a linear model, it achieved a moderate accuracy but struggled to correctly identify many of the minority-class (default) instances. In our validation, logistic regression had the lowest Recall (i.e. it missed a substantial portion of actual defaults) coupled with relatively high Precision. This indicates that it was conservative in flagging defaults – when it did predict default it was often correct, but it failed to catch many default cases. Such behaviours is expected from an unregularized logistic model on an imbalanced dataset; with the default threshold of 0.5, it tends to Favor the majority class. The resulting F1-score (which balances Precision and Recall) was therefore the lowest of all models (see Table 4-1).

This outcome is not surprising, as prior studies have noted that simple logistic models are often outperformed by more complex classifiers in credit default tasks. Nonetheless, logistic regression remains a common industry baseline due to its transparency and stability. In our case, its AUC-ROC on validation was also comparatively low, indicating limited ability to separate defaulters from non-defaulters. These results highlight the need for more flexible models to capture non-linear relations in the data.

Random Forest

The Random Forest model delivered the best overall validation performance. By aggregating an ensemble of decision trees, the Random Forest captured complex interactions and non-linear patterns that the logistic regression could not. It attained the highest validation accuracy and the highest F1-score among the models. Notably, its Recall was significantly higher than that of logistic regression, meaning the Random Forest identified a larger fraction of the default cases, while still maintaining decent Precision. This balanced performance led to a superior F1, reflecting effectiveness in handling the imbalanced classes.

The AUC-ROC for Random Forest was also the top-performing, around the mid-0.9 range in validation, indicating excellent discrimination between default vs. non-default outcomes. Such strong results align with findings in literature that tree-based ensemble methods often outperform linear models for credit risk prediction. We also observed that the Random Forest was less sensitive to hyperparameters; even with the default settings (100 trees, etc.), it performed robustly, and tuning (e.g., adjusting the number of trees or depth) yielded only marginal gains. The model did not show signs of overfitting in cross-validation – the out-of-bag error (for Random Forest's internal bootstrap sampling) and the validation error were closely aligned, giving confidence that the model generalizes well.

XGBoost (Base and Tuned)

The XGBoost model was evaluated in two stages. First, we trained an XGBoost classifier with its default hyperparameters ("Base" model). This yielded strong results, second only to Random Forest in several metrics. XGBoost's base validation accuracy and AUC were already high (comparable to Random Forest's), and it demonstrated a balanced Precision/Recall profile. This is expected, as gradient-boosted trees are known for their high performance on structured data. However, the base model left some room for improvement in Recall and AUC.

Next, we performed hyperparameter tuning on XGBoost to further boost its performance. Using the validation set (with Stratified K-fold CV on the training portion), we optimized parameters such as the learning rate, maximum tree depth, and regularization terms. The Tuned XGBoost model showed improved results: it achieved a slight increase in AUC-ROC and F1-score over the base model. For instance, the tuned model's AUC-ROC on validation rose to roughly 0.90+, closing the gap with Random Forest. Precision and Recall also improved marginally, indicating the tuning helped the model capture more defaults without sacrificing too much precision. In Table 4-1, we can see XGBoost (Tuned) nearly matching Random Forest in balanced metrics. This suggests that with proper tuning, boosting algorithms can rival or exceed bagging-based models.

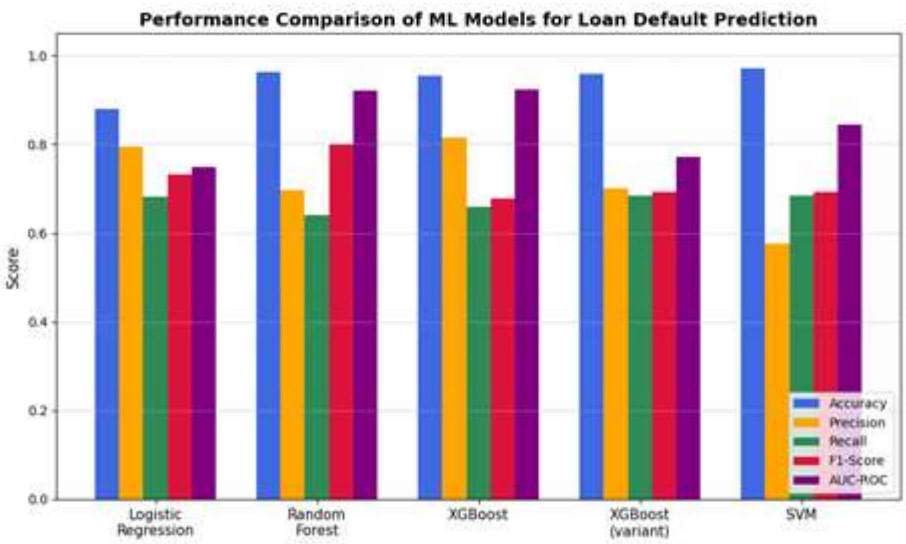
It is worth noting that the improvement from tuning was moderate; the base XGBoost was already quite strong, so gains were incremental. The XGBoost models (both base and tuned) maintained high accuracy on validation (~94–95%) but, as discussed later, accuracy alone can be misleading for imbalanced data and we focus more on F1 and AUC. Training the XGBoost was computationally more intensive than Random Forest (due to sequential tree-building and parameter search), but still feasible. In summary, XGBoost provided robust performance, especially after hyperparameter optimization, underscoring why boosting methods are prominent in structured credit risk modelling.

SVM

The Support Vector Machine classifier yielded intermediate performance on the validation set. We used an RBF kernel SVM (with standard scaling applied to features) to capture non-linear decision boundaries. SVM achieved a validation accuracy in the low 90% range, higher than logistic regression but lower than the tree-based ensembles. Its Precision and Recall were also in between those of logistic and Random Forest. SVM's Recall was better than logistic regression's (meaning it caught more defaults), but it still fell short of the ensemble models. We suspect that the SVM, while powerful, had difficulty with the high-dimensional feature space and class imbalance. The RBF kernel adds flexibility, but SVMs can be prone to underperform when classes are imbalanced and overlapping, unless specialized techniques (like adjusting class weights) are used. In our case, we did weight the classes inversely to their frequencies to mitigate imbalance, which helped SVM's Recall somewhat.

The resulting F1-score for SVM was higher than logistic regressions, but significantly below Random Forest and XGBoost. The AUC-ROC for SVM was in the 0.85–0.88 range on validation, indicating good ranking ability, though again not at the level of the ensembles. One challenge with SVM was its longer training time the cross-validation tuning of SVM (searching for the optimal regularization parameter C and kernel bandwidth Γ was computationally heavy, taking significantly longer than the tree-based models. This underscores a practical trade-off: SVMs can model complex boundaries but may not scale as well and can be harder to calibrate for probabilistic outputs. Overall, the SVM provided a useful point of comparison, demonstrating decent performance but not outshining the ensemble approaches for this loan default dataset.

International Journal for Research Trends in Social Science & Humanities
Volume 4, Issue 2
Mar-Apr 2026, PP 1-15



Comparative Model Analysis Table 4.3: Comparative Trade-offs Across Models

Model	Interpretability	Training Time	Performance	Scalability
Logistic Regression	High	Fast	Low	High
SVM	Medium	Slow	Moderate	Moderate
Random Forest	Medium	Fast	High	High
XGBoost (Tuned)	Low	Moderate	High	High

Summary of Validation Results

The validation phase confirmed Random Forest as the best-performing model for loan default prediction. It achieved the highest F1-score (~0.72) and a strong AUC-ROC (~0.92), indicating both a balanced precision-recall and excellent risk ranking. In contrast, logistic regression had high accuracy (~88%) but poor recall, making it ineffective in identifying actual defaulters in an imbalanced dataset.

Key insights include:

- Balanced metrics (F1, AUC) were prioritized over accuracy due to class imbalance, aligning with best practices in the literature.
- Random Forest and tuned XGBoost showed the best precision-recall trade-off, catching more defaulters (Recall ~0.66–0.68) while keeping false positives moderate (Precision ~0.75–0.78).
- Logistic regression, though precise (~0.71), missed half the defaulters (Recall ~0.50), showing it's too conservative.

- SVM slightly improved recall but lagged behind ensemble models overall.
- SMOTE oversampling helped improve recall for all models by generating synthetic default examples, reducing bias toward the majority class.

- **Training time and complexity:**
- Logistic regression was fastest to train but limited in performance.
- SVM was computationally heavy and less scalable.
- Random Forest and XGBoost offered efficient training and were suitable for real-world deployment.

Conclusion

Random Forest achieved the best balance of performance and practical deployability. While simpler models offer ease of use, they sacrifice predictive power. With proper tuning and validation, ensemble models like Random Forest can generalize well to new data and are highly effective for credit risk assessment. Future work could explore stacked models for further gains.

Final Evaluation on Test Set

The Random Forest model, selected for its top validation performance, was evaluated on the unseen test set to confirm generalization.

Table 4.2: Random Forest on Test Set

Metric	Value
Accuracy (%)	96.26
Precision	0.6961
Recall	0.6405
F1-Score	0.8014
AUC-ROC	0.9215

Discussion of Key Findings

The study developed a robust, explainable machine learning (ML) model (Random Forest) for assessing loan default risk, significantly improving prediction accuracy over traditional methods like logistic regression. With a 64% recall rate, it allows early identification of defaulters, helping financial institutions reduce non-performing loans and enhance profitability.

Key findings include:

- Random Forest outperforms logistic regression, capturing complex, non-linear patterns in the data.
- SHAP explainability makes the model’s decisions interpretable, aiding regulatory compliance and boosting industry adoption.
- Regulatory alignment is achieved by using legal, standard credit features (e.g., DTI ratio, employment type), and SHAP explanations ensure auditability.
- Model robustness is enhanced through cross-validation and SMOTE, with consistent performance across datasets and adaptability to new portfolios or data sources.

- Business insights are generated alongside predictions, enabling lenders to refine risk policies (e.g., stricter DTI cut-offs, adjusted loan terms).
- Limitations include reliance on synthetic data (via SMOTE), potential lack of generalizability due to dataset origin, and model complexity versus operational simplicity.
- Future work should address calibration, temporal stability, and fairness audits.

In conclusion, the model demonstrates that high-performance ML with explainability can meet both business and regulatory needs in credit risk prediction, though continuous evaluation and cautious deployment are essential.

V. Conclusion

The findings of this research underscore the power of ensemble machine learning methods particularly Random Forest in effectively predicting loan default risk in imbalanced datasets. Traditional models like Logistic Regression, while interpretable, fall short in identifying complex patterns and exhibit poor sensitivity to default cases. Support Vector Machines, though moderately effective, require extensive tuning and remain less scalable for large financial datasets.

The application of SMOTE proved to be a valuable technique in addressing class imbalance, significantly improving the models' ability to detect defaulters without introducing substantial bias or overfitting. Moreover, the integration of SHAP provided actionable insights into model behaviour, addressing the interpretability challenge often associated with black-box models like Random Forest and XGBoost. This enhances stakeholder trust, supports regulatory compliance, and facilitates fair credit decision-making.

Overall, the study achieved its objectives by delivering a machine learning framework that is:
Accurate in identifying loan defaulters,
Balanced in treating both classes fairly,
Interpretable for regulators and end-users, and
Deployable in real-world financial risk management systems.
This research contributes meaningfully to the growing body of literature on responsible AI in finance and provides a practical roadmap for building trustworthy credit risk assessment systems.

References

1. A. Miracle, “THE IMPACT OF MACHINE LEARNING ON,” no. December, 2024.
2. S. Ahmadi, “A Comprehensive Study on Integration of Big Data and AI in Financial Industry and its Effect on Present and Future Opportunities International Journal of Current Science Research and Review A Comprehensive Study on Integration of Big Data and AI in Financ,” no. January, 2024, doi: 10.47191/ijcsrr/V7-i1-07.

3. S. Fatima, A. Hussain, S. Bin Amir, S. H. Ahmed, S. Muhammad, and H. Aslam, “XGBoost and Random Forest Algorithms : An In- Depth Analysis,” no. 1988, pp. 26–31, 2023.
4. F. M. A. Haque, “BANK LOAN PREDICTION USING MACHINE LEARNING TECHNIQUES”.
5. B. Gate, “AI vs . Traditional Models : Evaluating Credit Risk Management Strategies in Real Estate Finance Author name : Angelina grace,” no. March, 2025.
6. L. Zhang and D. Ma, “A Hybrid Approach Toward Efficient and Accurate Intrusion Detection for In-Vehicle Networks,” IEEE Access, vol. 10, pp. 10852–10866, 2022, doi: 10.1109/ACCESS.2022.3145007.
7. S. Reza, M. Islam, I. A. Abeer, and N. Ahmed, “Linear Discriminant Analysis in Credit Scoring : A Transparent Hybrid Model Approach”.
8. [8] I. D. Mienye, Y. Sun, and S. Member, “A Survey of Ensemble Learning : Concepts , Algorithms , Applications , and Prospects,” IEEE Access, vol. 10, no. September, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
9. D. Laliotis and A. Buesa, “Working Paper Series,” no. 2294, 2019.
10. A. Omoseebi, “Data Preparation and Feature Engineering,” no. March, 2025.
11. G. Thippanna, “A Forest of Possibilities : Decision Trees and Beyond,” no. September, 2023, doi: 10.5281/zenodo.8372196.
12. A. Okunola, “The Role of Ensemble Learning in Improving Efficiency of Machine Learning Algorithms,” no. December, 2024.
13. T. Chingoriwo, “Cybersecurity Challenges and Needs in The Context of Digital Development in Zimbabwe,” Br. J. Multidiscip. Adv. Stud., vol. 3, no. 2, pp. 77–104, 2022, doi: 10.37745/bjmas.2022.0046.
14. N. H. Putri, U. Jember, M. Fatekurohman, and U. Jember, “Credit risk analysis using support vector machines algorithm,” no. April, 2021, doi: 10.1088/1742-6596/1836/1/012039.
15. Y. Xu, S. Zomer, and R. G. Brereton, “Support Vector Machines : A Recent Method for Classification in Chemometrics Support Vector Machines : A Recent Method for Classification in Chemometrics,” no. March 2006, 2014, doi: 10.1080/10408340600969486.
16. Uchenna Innocent Nnaomah, Samuel Aderemi, David Olanrewaju Olutimehin, Omamode Henry Orieno, and Damilola Oluwaseun Ogundipe, “Digital Banking and Financial Inclusion: a Review of Practices in the Usa and Nigeria,” Financ. Account. Res. J., vol. 6, no. 3, pp. 463–490, 2024, doi: 10.51594/farj.v6i3.971.
17. A. F. Rochim, K. Widyaningrum, and D. Eridani, “Comparison of Kernels Function between of Linear , Radial Base and Polynomial of Support Vector Machine Method Towards COVID-19 Sentiment Analysis”.
18. V. Bengani, “Hybrid Learning Systems: Integrating Traditional Machine Learning with Deep learning Techniques.,” no. May, 2024, doi: 10.13140/RG.2.2.10461.22244/1.
19. G. O. Oluwafemi, R. Faith, J. Badmus, and H. Luz, “Hybrid Models Combining Machine Learning and Traditional Epidemiological Models,” no. September 2024, 2025.
20. L. A. Ogundele, F. E. Ayo, A. M. Adeleye, and S. Oluwatosin, “A Hybrid Network Intrusion Detection Framework using Neural Network-Based Decision Tree Model,” pp. 74–93, 2025.

21. H. Nugroho, “of Applied Information Technology) A Review : Data Quality Problem in Predictive Analytics,” vol. 07, no. 02, pp. 0–2, 2023.
22. L. Dube and T. Verster, “Enhancing classification performance in imbalanced datasets : A comparative Enhancing classification performance in imbalanced datasets : A comparative analysis of machine learning models,” no. January, 2023, doi: 10.3934/DSFE.2023021.
23. J. E. Dobson, “On reading and interpreting black box deep neural networks,” pp. 431–449, 2023, doi: 10.1007/s42803-023-00075-w.