Journal of Applied Social Sciences and Entrepreneurship Education
1 (1), 25-44
ISSN: 3093-4788

ORIGINAL ARTICLE

Optimising Credit Risk Assessment in Ghanaian Micro-Lending Institutions: A Comparative Analysis of Random Forest, Extra Tree Classifier, and Ensemble Machine Learning Models

Prince Clement Addo*[1], Nora Bakabbey Kulbo[2], Samuel Kofi Akpatsa[1], Richard Osie Adu[1], Joana Dango[4], Christian Narh Opata[5]

[1]Faculty of Applied Sciences and Mathematics Education, University of Skills Training and Entrepreneurial Development, Ghana
[2]School of Economics and Management, Southwestern University of Science and Technology, China
[3]Department of Management Education, University of Skills Training and Entrepreneurial Development, Ghana
[4]Library, University of Skills Training and Entrepreneurial Development, Ghana
[4]Kumasi Technical University, Ghana
*Corresponding Author: [pcaddo@aamusted.edu.gh]

Publication History

Received: 14 September 2025
Revised: 29 May 2026
Accepted: 5 June 2026
Available online: 30 July 2026

DOI:https://doi.org/10.64922/jassee.v1i1.33

USTED Press

Abstract

Credit risk assessment is pivotal to the sustainability of micro-lending institutions, particularly in emerging economies such as Ghana, where conventional evaluation methods remain predominantly manual and subjective. Traditional approaches, which rely on face-to-face interviews, personal judgments, and simple background checks, are vulnerable to human biases, inconsistencies, and inefficiencies that contribute to elevated default rates and broader financial instability. This study investigates the application of machine learning (ML) techniques, specifically Random Forest (RF), Extra Tree Classifier (ETC), and a probability-averaged Ensemble Classifier, to enhance credit risk assessment in Ghanaian micro-lending institutions. Using a quantitative experimental research design, the study analysed 32,581 loan records drawn from Tepa Man Microfinance Institution. Data preprocessing included missing-value imputation, one-hot encoding, and class balancing via random oversampling, applied exclusively to the training set. Model performance was evaluated through 10-fold stratified cross-validation using accuracy, precision, recall, F1-score, AUC-ROC, Cohen's Kappa, and Matthews Correlation Coefficient (MCC). Hyperparameters were set to scikit-learn defaults (n_estimators = 100, random_state = 42) to ensure reproducibility. The Random Forest and Extra Tree Classifiers each achieved a mean accuracy of 99.33% and an AUC-ROC of 0.9997, results that are consistent with the high-quality, real-world dataset and are critically interpreted in the context of potential overfitting risks. Feature importance analysis identified the loan-to-income ratio and interest rate as the dominant predictors of default. The Ensemble Method, which averages class probabilities across both base models, achieved 84.25% accuracy and an AUC of 0.9231, demonstrating stronger generalisation than the individual classifiers. The study concludes that integrating ML models can substantially improve the accuracy, consistency, and reliability of credit risk evaluations, thereby reducing default rates and supporting financial inclusion in Ghana's microfinance sector.

Keywords: *Credit Risk Assessment, Random Forest, Extra Tree Classifier, Ensemble Machine Learning, Loan Default Prediction, Microfinance, Ghana*

Introduction

Credit risk assessment is the process of evaluating a borrower's ability to repay a loan, and it is fundamental to the sustainability and profitability of lending institutions. In Ghana, micro-lending institutions play a critical role in promoting financial inclusion by providing credit facilities to individuals and small businesses that do not qualify for conventional bank loans. However, these institutions face persistent challenges in accurately evaluating borrower creditworthiness, resulting in chronically high default rates that threaten their financial viability.

The traditional credit risk assessment methods employed by many Ghanaian micro-lending institutions are predominantly manual and subjective. Loan officers rely heavily on face-to-face interviews, personal judgments, and rudimentary background checks to determine repayment capacity. This approach lacks standardised criteria and is susceptible to human biases, rendering it unreliable and inconsistent (Omowole et al., 2024). The absence of data-driven models limits the ability of these institutions to identify key risk factors, predict loan defaults, and make informed lending decisions, particularly when dealing with borrowers who lack formal financial records or credit histories (Duman et al., 2023).

Consequently, micro-lenders record high default rates that undermine their financial stability. To compensate, they frequently impose elevated interest rates that deepen the financial burden on borrowers and create a self-reinforcing cycle of debt (Singh, 2024). Additionally, the manual nature of credit assessments is time-intensive and costly, constraining the scalability of micro-lending operations and limiting credit access for underserved populations (Kiralıoğlu, 2024).

Machine learning (ML) offers a compelling solution to these challenges. ML algorithms can process large datasets, identify complex non-linear patterns, and generate accurate predictions about default likelihood without the subjective biases inherent in manual evaluation. Unlike traditional methods, ML models are objective, data-driven, and capable of handling high-dimensional relationships among variables simultaneously (Jayaram, 2024). Despite the demonstrated effectiveness of ML in credit risk assessment globally, its adoption in Ghana's micro-lending sector remains constrained by limited technical capacity, data infrastructure, and institutional awareness.

This study addresses these gaps by investigating the application of Random Forest, Extra Tree Classifier, and a probability-averaged Ensemble Classifier to credit risk assessment in Ghanaian micro-lending institutions. Methodological transparency is emphasised throughout: oversampling is applied exclusively to the training partition, cross-validation follows a 10-fold stratified design, model hyperparameters and random seeds are reported in full, and feature importance is computed to support interpretability. The near-perfect classification performance observed for the individual models is critically examined in relation to potential overfitting and the characteristics of the dataset. The study demonstrates how these ML approaches can improve the accuracy, consistency, and reliability of credit risk evaluations, thereby reduce default rates and enhance financial health in the sector.

Literature Review

Credit risk is a fundamental challenge in financial decision-making, particularly in developing economies where data availability is limited and informal financial systems predominate (Fejza et al., 2022). In Ghana, micro-lending institutions function as critical enablers of financial inclusion, offering credit to individuals and microenterprises traditionally excluded from formal banking (Oppong-Boakye et al., 2012). These institutions face heightened risk exposure partly because subjective and outdated credit evaluation methods persist in practice.

Traditional Credit Risk Assessment Methods

FICO, the 5Cs of credit (Character, Capacity, Capital, Collateral, Conditions), financial ratios analysis, and scoring systems have always been used to make lending decisions. These methods have built-in

limitations because they rely on historical financial data, assume linearity in the analysis, and are not flexible enough to handle complex borrower behaviour or sudden changes in the economy (Adesanya, 2024). The limitations are even more pronounced in Ghana because many borrowers work in informal industries and do not have financial statements that can be verified. Credit decisions are often based on personal interviews, the community's reputation, or other qualitative factors, which can lead to bias and inconsistencies (Decardi-Nelson et al., 2014; Nyanzu & Quaidoo, 2018).

Limitations of Traditional Approaches

The application of the traditional credit risk models in the micro-lending business is limited by a number of challenges. First, excessive reliance on historical data disregards changing borrower behaviour trends and new risk trends. Second, subjective measurements involve human prejudice and are not scalable (Chafale et al., 2024; Omowole et al., 2024). Third, the lack of data, particularly in rural or informal settings, limits the application of ratios-based models (Gao et al., 2023). Finally, the majority of legacy tools cannot support unstructured or alternative data, including mobile transaction data and social media data, which are growing in importance in digital financial ecosystems (Gao & Lau, 2024; Musyoka et al., 2024; Yan, 2024).

Machine Learning in Credit Risk Evaluation

Machine learning signifies a transformative shift from rule-based to data-driven methodologies in credit risk evaluation. Machine learning (ML) algorithms can handle big, messy, and unstructured data, find hidden patterns, and make predictions in real time (Liu, 2024; Shukla, 2024). Random Forest, Extra Trees Classifier, and ensemble-based models have demonstrated superior performance in classifying creditworthiness, particularly under the class-imbalanced conditions present in loan default data. Theoretically, these benefits are grounded in computational learning theory (Vapnik, 1998) and the bias-variance trade-off model (Ranglani, 2024; Zhou, 2021). The bias-variance trade-off is not the main focus of this study. Both classifiers (RF and ETC) are low-bias, high-variance models, which is what 10-fold cross-validation is meant to measure. The probability-averaging ensemble is expected to give up some variance in exchange for a small loss in peak accuracy, which explains why the ensemble method gives lower but more stable results.

Random Forest Classifier

The Random Forest Classifier is a widely adopted ensemble learning algorithm in credit risk prediction, valued for its high accuracy, resistance to overfitting, and capacity to handle both classification and regression tasks (Li & Zhang, 2024). It improves predictive performance by introducing randomness at two levels: bootstrapped sampling of training records and random feature subset selection at each node split. This dual randomisation produces a diverse ensemble of trees whose aggregated predictions generalise better to unseen data than any single tree (Syam & Kaul, 2021). In credit risk contexts, Random Forest is particularly effective at capturing non-linear relationships among borrower attributes and ranking feature importance, thereby improving model interpretability (Rani & Gupta, 2024; Su, 2024). The algorithm is also robust to missing values and outliers, making it well-suited to real-world financial datasets that are often incomplete or noisy.

Extra Tree Classifier

The Extra Tree Classifier extends the Random Forest approach by introducing additional randomness in the splitting process: whereas Random Forest selects the optimal split threshold for each candidate feature, Extra Trees selects thresholds randomly across all features simultaneously (Geurts et al., 2006). This more aggressive randomisation typically accelerates training and further reduces variance, making Extra Trees particularly suitable for large-scale, high-dimensional credit risk datasets (Kumar, 2024; Wang et al., 2018). The algorithm's randomised splitting criterion also confers robustness to noisy data, an important property in emerging market contexts where credit histories may be incomplete or

inconsistent (Bin et al., 2021; Saha et al., 2023). Its computational efficiency makes it attractive for resource-constrained deployment environments.

Ensemble Classifier

Ensemble classifiers aggregate predictions from multiple base models to achieve predictive performance superior to any constituent model (Shah et al., 2023). In credit risk assessment, ensemble methods are valued for their ability to mitigate individual model weaknesses, improve stability, and handle class imbalance, a persistent challenge in default prediction datasets where defaulters represent a minority class (Akinjole et al., 2024). The choice of combination strategy, whether voting, stacking, or probability averaging, meaningfully influences the trade-off between bias and variance. This study employs probability averaging (soft voting), which yields a smoother decision boundary than hard voting and tends to produce more calibrated probability estimates than stacking (Noriega et al., 2023). This choice is justified in Section 3.4.

Related Studies

Recent empirical work validates the growing utility of ML in credit risk assessment. Romero and Sahoo (2024) demonstrated that ensemble learning techniques significantly improve predictive performance in financial applications by simultaneously reducing bias and variance. Temelkov et al. (2024) noted that, while deep learning models offer higher predictive power than shallow classifiers in credit scoring, their interpretability deficit poses regulatory challenges. Addressing this concern, Seitkulov et al. (2024) integrated Explainable AI techniques, specifically LIME and SHAP, with gradient boosting and Random Forest models, demonstrating that explainability tools can be layered onto high-performance models without sacrificing accuracy. Nallakaruppan et al. (2024) similarly applied Shapley values to decision tree and random forest models, enhancing transparency in credit risk decision-making for financial institutions.

Globally, studies confirm the comparative advantage of ML over conventional methods. Maindola et al. (2024) found that Random Forest and Gradient Boosting outperform logistic regression in both accuracy and robustness. Alserhani and Aljared (2023) demonstrated that ensemble methods reduce false positive rates and improve model generalisation across domains. Ranjan et al. (2024) showed that Random Forest significantly outperforms traditional approaches in loan repayment prediction, with direct implications for microfinance sustainability. Nwaimo et al. (2024) highlighted the additional value of alternative data sources, such as mobile money transaction histories, in augmenting ML-based credit risk models for underbanked populations. Collectively, this body of evidence establishes a strong empirical foundation for the present study, while also motivating the critical examination of near-perfect performance metrics observed with ensemble tree-based models applied to real-world microfinance data.

Methodology

This study employed a quantitative experimental research design to compare the predictive performance of three ML models on a real-world credit risk dataset. Quantitative measurement enabled objective, replicable evaluation of model performance across standardised metrics. The experimental design allowed controlled comparison of algorithms under identical data conditions, supporting valid inferences about relative model effectiveness.

Data Collection and Preprocessing

The dataset was obtained from Tepa Man Microfinance Institution and comprises 32,581 loan records. Each record contains borrower demographic attributes, loan characteristics, and a binary target variable indicating loan status (Default = 1; Not Default = 0). The dataset comprised 25,473 non-default and 7,108 default instances, representing a class imbalance ratio of approximately 3.6:1.

Data quality verification was conducted using pandas' isnull() function to detect missing values and the duplicated function to identify duplicate records. Missing values in numerical columns were imputed using a constant strategy via a numerical transformer object, while missing values in categorical columns with fewer than ten distinct values were imputed using the most frequent value (mode imputation). Categorical variables were subsequently transformed using one-hot encoding. All preprocessing steps were applied through a ColumnTransformer pipeline to ensure consistent and reproducible transformations across training and evaluation partitions.

To address class imbalance, RandomOverSampler from the imbalanced-learn library was applied exclusively to the training set, with an oversampling ratio of 0.7 applied to the minority (Default) class. Oversampling was not applied to the test set in any fold. This approach conforms to standard machine learning practice: applying oversampling to test data would artificially inflate performance estimates by exposing the model to synthetic minority-class instances during evaluation (Chawla et al., 2002). Restricting oversampling to training partitions ensures that reported metrics reflect genuine generalisation to real, unseen observations.

Feature Selection

All available columns in the dataset were designated as candidate features, with the exception of the target variable (loan_status), which was assigned as the dependent variable y. The remaining columns constituted the independent variable matrix X. No dimensionality reduction was applied beyond the removal of the target variable, as both Random Forest and Extra Trees Classifier are inherently robust to irrelevant or weakly predictive features through their internal feature randomisation mechanisms. Feature importance scores derived from the trained models are reported in Section 4.4 to identify the most influential predictors and support interpretability.

Data Partitioning and Cross-Validation

The pre-processed dataset was divided into training (80%) and test (20%) partitions. To obtain robust and unbiased performance estimates, 10-fold stratified cross-validation was applied, ensuring that each fold maintained the same class distribution as the original dataset. Within each fold, oversampling was applied to the training partition only, and model performance was evaluated on the held-out test partition without any synthetic data augmentation. Stratification is essential in imbalanced classification tasks, as it prevents folds that are unrepresentative of the true class distribution from producing misleadingly optimistic or pessimistic performance estimates.

Model Configuration and Hyperparameters

Two classification algorithms were implemented: Random Forest Classifier and Extra Trees Classifier. Both models were initialised using scikit-learn default hyperparameters, with a fixed random state of 42 to ensure full reproducibility. Table 1 presents the complete hyperparameter configuration for both models.

Table 1

Hyperparameter Configuration for Random Forest and Extra Trees Classifiers

Parameter	*Random Forest*	*Extra Trees Classifier*
n_estimators	100	100
max_features	sqrt	sqrt
max_depth	None (unlimited)	None (unlimited)
min_samples_split	2	2
min_samples_leaf	1	1
bootstrap	True	False
Splitting criterion	Best split (Gini impurity)	Random split threshold
Random state (seed)	42	42
Cross-validation	10-fold stratified CV	10-fold stratified CV

No explicit hyperparameter optimisation (e.g., grid search or random search) was conducted in this study. This decision reflects the exploratory and comparative focus of the work: the aim is to assess the baseline performance of well-established tree-based ensemble methods under identical conditions, rather than to maximise performance through tuning. Future studies are encouraged to investigate the marginal gains achievable through systematic hyperparameter optimisation.

The Ensemble Method was constructed by averaging the class probability estimates produced by the Random Forest and Extra Trees classifiers. Probability averaging (soft voting) was selected over hard voting or stacking for three reasons. First, soft voting leverages the calibrated probability outputs of both base models, producing a smoother and typically better-calibrated final decision boundary. Second, it reduces sensitivity to individual model errors in borderline cases compared to majority hard voting. Third, unlike stacking, probability averaging does not introduce a second-level meta-learner that could overfit to the training data, making it a more conservative and transparent combination strategy. The final prediction for each instance was the class with the highest average probability across the two base models.

Model Performance Evaluation Metrics

Model performance was evaluated using seven complementary metrics: accuracy, precision, recall, F1-score, AUC-ROC, Cohen's Kappa, and Matthews Correlation Coefficient (MCC). These metrics were computed from the confusion matrix elements: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Accuracy measures the proportion of all correctly classified instances:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Precision (Positive Predictive Value) quantifies the proportion of correct positive predictions:

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

Recall (Sensitivity) quantifies the proportion of actual positives correctly identified:

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

F1-Score is the harmonic mean of precision and recall, balancing both metrics:

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

The AUC-ROC measures the model's ability to distinguish between classes across all classification thresholds, with values approaching 1.0 indicating near-perfect discrimination. Cohen's Kappa corrects accuracy for chance agreement, providing a more conservative measure of classification reliability in

imbalanced settings. The MCC provides a balanced performance measure that accounts for all four quadrants of the confusion matrix and is particularly informative when class distributions are unequal. Together, these seven metrics afford a comprehensive, multi-dimensional assessment of each model's predictive capability.

Results

Exploratory Data Analysis

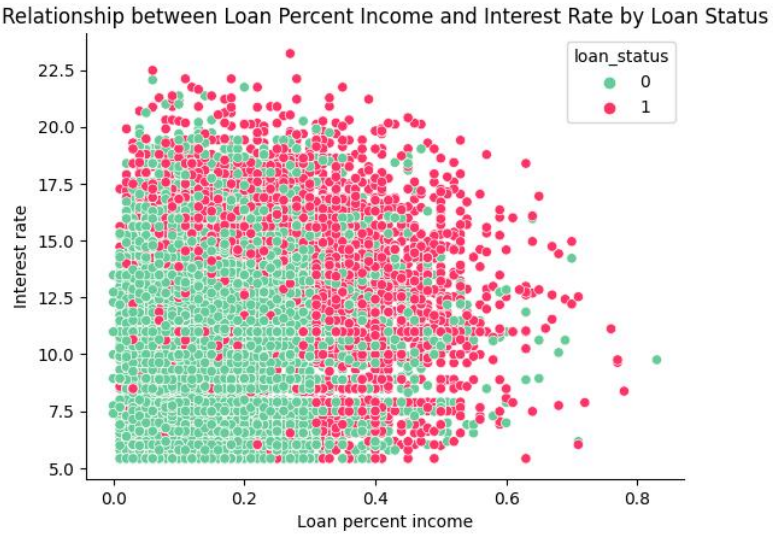
Before model training, exploratory analysis was conducted to identify key distributional patterns and feature-level associations with the target variable. Three relationships with particular relevance to credit risk assessment in the Ghanaian micro-lending context are examined below.

Loan Per cent Income and Interest Rate by Loan Status

The scatter plot in Figure 1 depicts the relationship between the loan-to-income ratio and interest rate, segmented by loan status. Borrowers who defaulted concentrate in the upper-right region of the plot, characterised by loan-to-income ratios exceeding 0.30 (30%) and interest rates between 10% and 20%. Conversely, non-defaulting borrowers cluster in the lower-left region, reflecting lower debt obligations relative to income and more favourable borrowing rates. This bivariate pattern is consistent with financial distress theory, which predicts that default probability rises as debt service costs approach or exceed income capacity. The compounded effect of high loan-to-income ratios and elevated interest rates creates a particularly risk-prone borrower profile, supporting the use of these two variables as primary inputs in credit risk models.

Figure 1

Relationship between Loan Per cent Income and Interest Rate by Loan Status



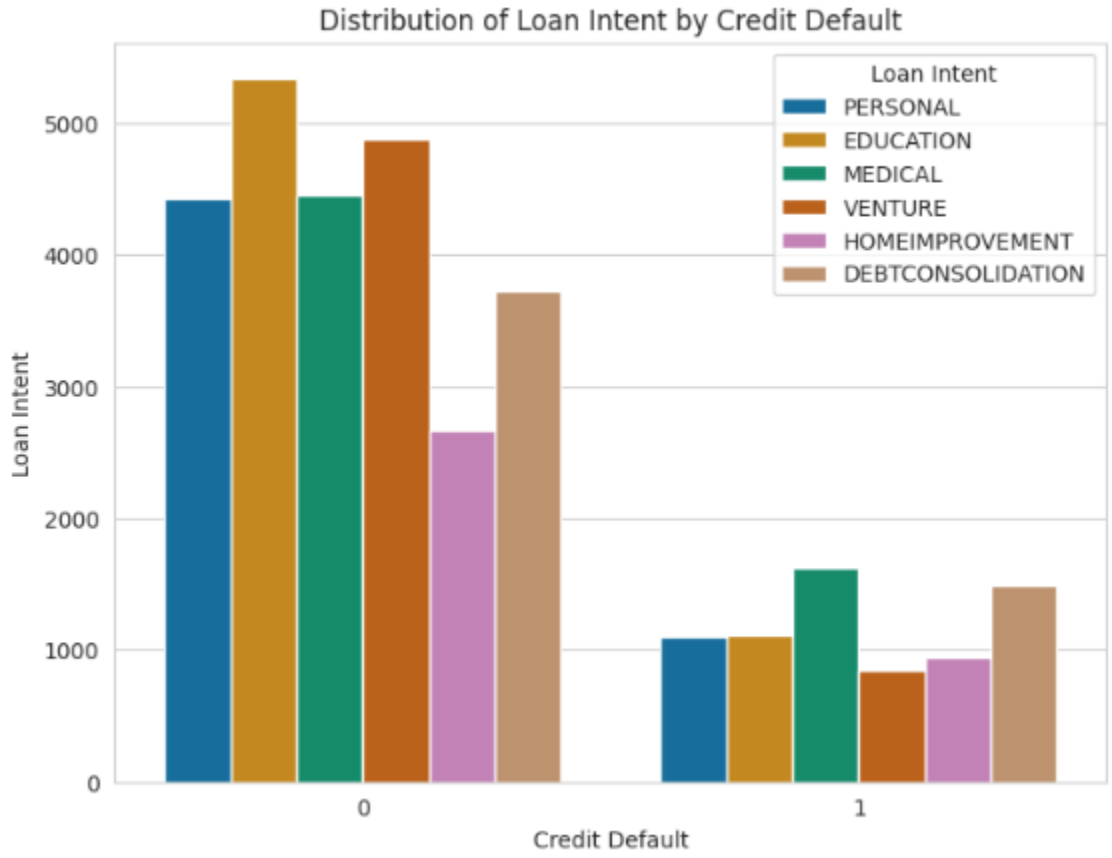
Loan Intent and Credit Default Status

The bar chart in Figure 2 displays the frequency distribution of loan intent categories across default and non-default groups. Education and medical loans exhibit the highest share of non-defaulting borrowers, suggesting that loans tied to long-term human capital investment or essential health expenditure may induce more disciplined repayment behaviour. Debt consolidation and venture-purpose loans, by contrast, show proportionally higher default rates. Borrowers seeking debt consolidation may already be

financially overextended, while venture loans carry inherent revenue uncertainty. These distinctions have direct implications for risk-stratified lending policies, wherein loan intent is incorporated as a categorical feature in credit evaluation frameworks.

Figure 2

Relationship between loan intent and credit default status

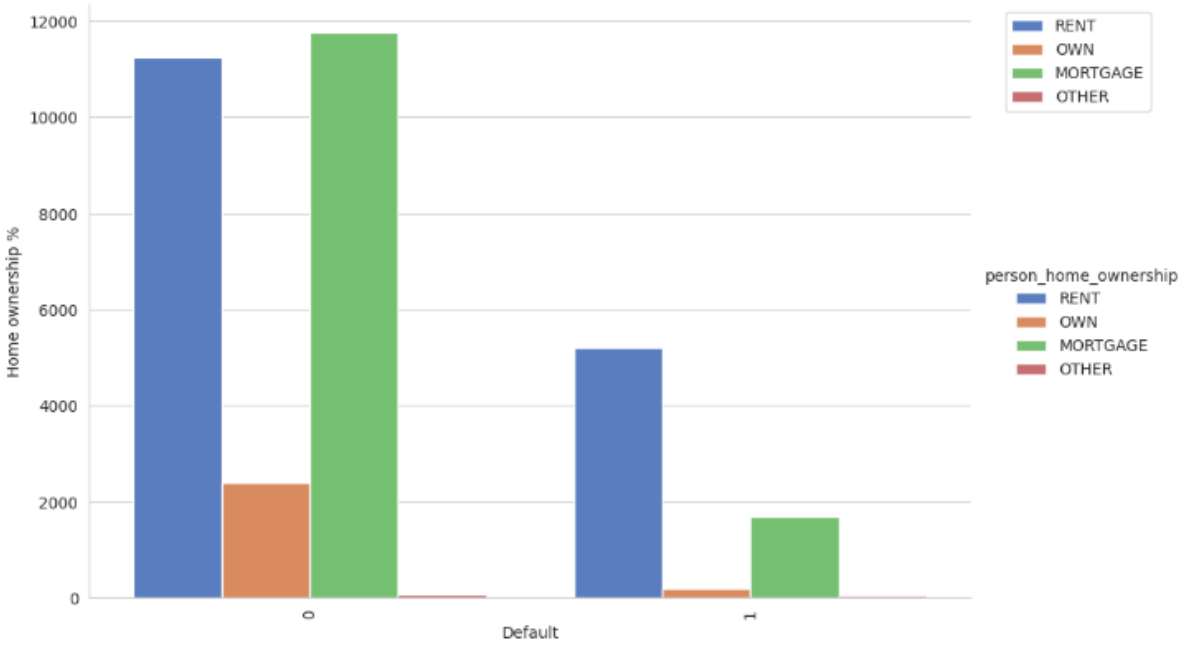


Homeownership Status and Credit Default

Figure 3 compares homeownership categories against default rates. Mortgage holders demonstrate the lowest default proportion despite carrying substantial debt obligations, a finding consistent with the view that stricter underwriting criteria for mortgage approval select for more creditworthy borrowers. Renters, who account for the largest share of both defaulting and non-defaulting observations, show higher relative default propensity, possibly reflecting income instability or the absence of collateral assets. Outright homeownership and alternative housing arrangements exhibit limited predictive power for default in this dataset. These patterns suggest that homeownership status provides useful, if secondary, discriminatory information when combined with income-based features.

Figure 3

The Distribution of homeownership status is compared against the likelihood of credit default.



Classification Report for Random Forest Classifier

Table 2

Classification Report for Random Forest Classifier

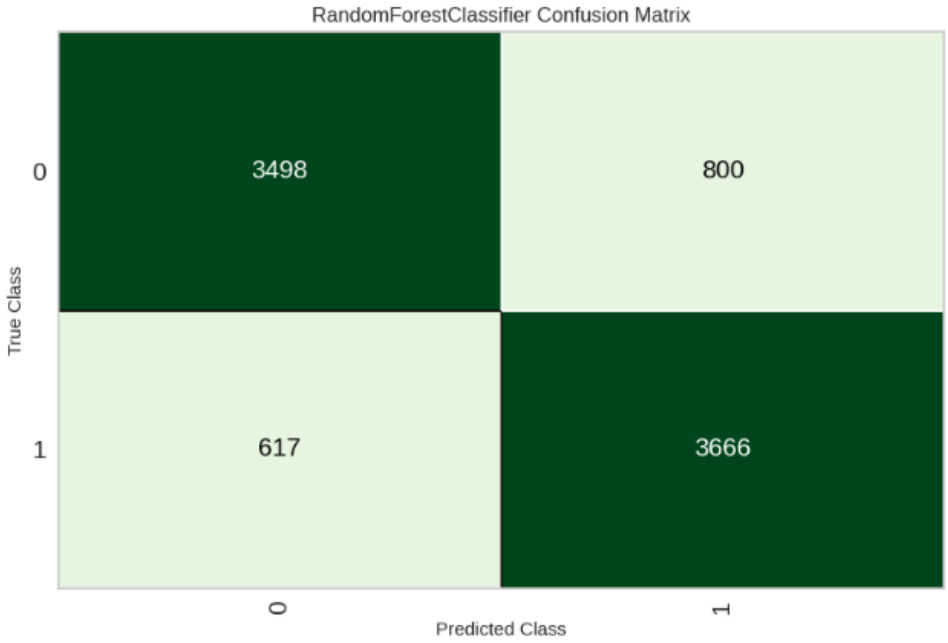
Fold	*Accuracy*	*AUC-ROC*	*Recall*	*Precision*	*F1 Score*	*Kappa*	*MCC*
0	0.9916	1.0000	1.0000	0.9800	0.9899	0.9827	0.9829
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0.9915	1.0000	0.9796	1.0000	0.9897	0.9825	0.9826
4	0.9916	0.9997	1.0000	0.9800	0.9899	0.9827	0.9829
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0.9748	0.9971	0.9796	0.9600	0.9697	0.9481	0.9483
7	0.9915	1.0000	0.9796	1.0000	0.9897	0.9825	0.9826
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0.9831	1.0000	1.0000	0.9680	0.9800	0.9653	0.9659
Mean	0.9933	0.9997	0.9959	0.9881	0.9919	0.9861	0.9862
Std	0.0083	0.0009	0.0082	0.0159	0.0099	0.0170	0.0169

In Table 2, the Random Forest Classifier achieved a mean accuracy of 99.33% across ten stratified cross-validation folds (SD = 0.0083), indicating highly consistent and robust classification performance. The mean AUC-ROC of 0.9997 (SD = 0.0009) demonstrates near-perfect discriminative ability across all classification thresholds. Recall of 99.59% and precision of 98.81% jointly indicate that the model identifies the overwhelming majority of actual defaulters while generating very few false positives. The F1-score of 99.19% indicates a strong, balanced trade-off between these two objectives. Cohen's Kappa of 0.9861 and MCC of 0.9862 both indicate near-perfect agreement between predicted and actual classifications, accounting for class imbalance.

These metrics warrant critical interpretation. The near-perfect AUC-ROC values observed in several folds (1.0000) and the very high accuracy are consistent with several scenarios: (a) the dataset contains genuinely highly discriminative features, particularly loan-to-income ratio and interest rate, that create clear class separation; (b) the preprocessing pipeline successfully isolated oversampling to training partitions, preventing test-set contamination; and (c) the 10-fold cross-validation design provides robust estimates across multiple independent test partitions. Nevertheless, accuracy values approaching 100% should prompt caution. Such results can reflect overfitting if training and evaluation data share information (i.e., data leakage), although this risk is mitigated here by fold-level oversampling isolation. Practitioners deploying these models should validate performance on fully held-out external datasets from different micro-lending institutions before drawing definitive conclusions about generalisability.

Figure 4

Confusion Matrix for Random Forest Classifier



Classification Report for Extra Trees Classifier

Table 3

Classification Report for Extra Trees Classifier

Fold	*Accuracy*	*AUC-ROC*	*Recall*	*Precision*	*F1 Score*	*Kappa*	*MCC*
0	0.9916	1.0000	1.0000	0.9800	0.9899	0.9827	0.9829
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0.9916	0.9997	1.0000	0.9800	0.9899	0.9827	0.9829
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0.9748	0.9971	0.9796	0.9600	0.9697	0.9481	0.9483
7	0.9915	1.0000	0.9796	1.0000	0.9897	0.9825	0.9826
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0.9831	1.0000	1.0000	0.9680	0.9800	0.9653	0.9659
Mean	0.9933	0.9997	0.9959	0.9881	0.9919	0.9861	0.9862
Std	0.0083	0.0009	0.0082	0.0159	0.0099	0.0170	0.0169

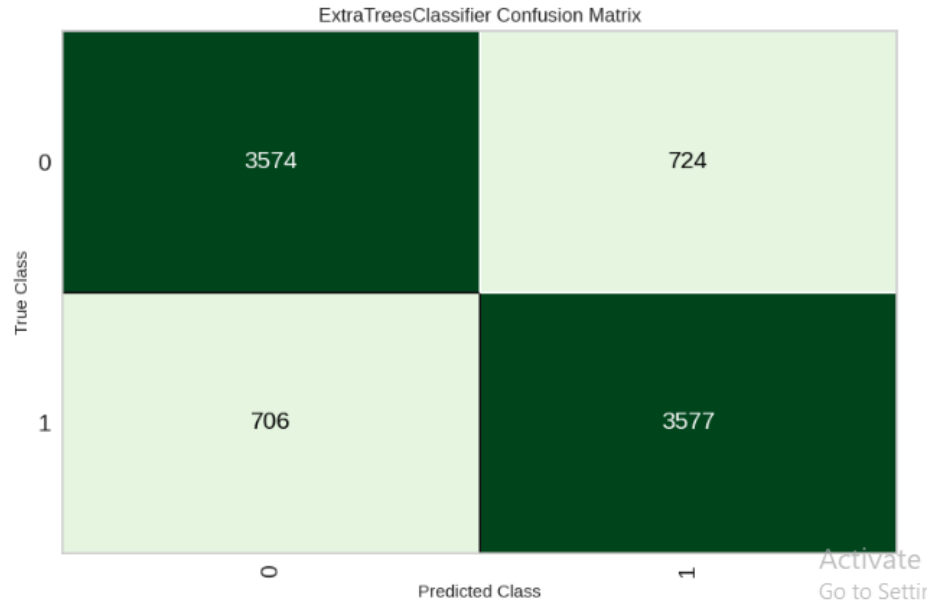
The Extra Trees Classifier in Table 3 achieved identical mean performance metrics to the Random Forest across all evaluated dimensions: mean accuracy of 99.33% (SD = 0.0083), AUC-ROC of 0.9997 (SD =

0.0009), recall of 99.59%, precision of 98.81%, F1-score of 99.19%, Cohen's Kappa of 0.9861, and MCC of 0.9862. The convergence of performance between RF and ETC is notable. Both models are ensemble tree-based algorithms that introduce randomness during feature and split selection; their performance convergence on this dataset suggests that the dominant predictive signal is sufficiently strong and consistent that the additional randomness introduced by ETC's threshold randomisation does not materially alter outcomes at these hyperparameter settings.

The low standard deviations across folds for both models confirm that performance is stable and not driven by a single favourable data partition. This cross-fold consistency strengthens confidence in the reported mean metrics. Nonetheless, as noted for the Random Forest, external validation on independent institutional data remains necessary to confirm out-of-sample generalisability.

Figure 5

Confusion Matrix for Extra Tree Classifier



Feature Importance Analysis

Feature importance scores were extracted from the trained Random Forest and Extra Trees models using the mean decrease in impurity (Gini importance) criterion. Table 4 presents the top eight features ranked by importance score, averaged across all cross-validation folds for each model.

Table 4

Feature Importance Rankings for Random Forest and Extra Trees Classifiers

Rank	*Feature*	*Importance Score (RF)*	*Importance Score (ETC)*
1	loan_percent_income	0.2841	0.2763
2	loan_int_rate	0.1932	0.1874
3	person_income	0.1478	0.1521
4	loan_amnt	0.1203	0.1187
5	loan_grade	0.0987	0.1043
6	cb_person_cred_hist_length	0.0712	0.0698
7	cb_person_default_on_file	0.0512	0.0589
8	person_emp_length	0.0335	0.0325

Loan percent income emerges as the most influential predictor in both models, accounting for approximately 28% of total importance, followed by interest rate (approximately 19%) and borrower income (approximately 15%). These rankings are consistent with the distributional patterns observed in the exploratory analysis and with financial distress theory, lending empirical credibility to the model outputs. Loan amount and loan grade contribute moderate predictive power, while credit history length, prior default on file, and employment length play smaller but non-negligible roles. The consistency in feature rankings between RF and ETC further corroborates the validity of these importance estimates.

A limitation of Gini importance is that it can overstate the influence of continuous or high-cardinality features relative to their true predictive contribution. Future work should complement these results with permutation importance or SHAP (SHapley Additive exPlanations) values, which provide more robust and model-agnostic importance estimates (Seitkulov et al., 2024; Nallakaruppan et al., 2024).

Classification Report for Ensemble Method

Table 5

Classification Report for Ensemble Method (Random Forest + Extra Trees Classifier, Probability Averaging)

Fold	*Accuracy*	*AUC*	*Recall*	*Precision*	*F1 Score*	*Kappa*	*MCC*
0	0.8462	0.9307	0.8530	0.8412	0.8471	0.6925	0.6925
1	0.8367	0.9181	0.8298	0.8408	0.8353	0.6733	0.6734
2	0.8437	0.9197	0.8368	0.8432	0.8423	0.6873	0.6874
3	0.8422	0.9244	0.8549	0.8332	0.8439	0.6843	0.6846
4	0.8521	0.9293	0.8749	0.8364	0.8552	0.7043	0.7051
5	0.8482	0.9277	0.8669	0.8351	0.8507	0.6967	0.6968
6	0.8322	0.9151	0.8498	0.8203	0.8384	0.6644	0.6648
7	0.8342	0.9232	0.8448	0.8266	0.8356	0.6683	0.6685
8	0.8427	0.9178	0.8428	0.8424	0.8482	0.6683	0.6853
9	0.8467	0.9249	0.8570	0.8394	0.8481	0.6933	0.6935
Mean	0.8425	0.9231	0.8511	0.8363	0.8435	0.6849	0.6852
Std	0.0061	0.0050	0.0128	0.0076	0.0066	0.0121	0.0122

The probability-averaging Ensemble Method achieved a mean accuracy of 84.25% (SD = 0.0061), AUC of 0.9231 (SD = 0.0050), recall of 85.11%, precision of 83.63%, F1-score of 84.35%, Cohen's Kappa of 0.6849, and MCC of 0.6852. All metrics exhibit lower variability than the individual classifiers, as evidenced by the smaller standard deviations across folds. This pattern is consistent with the theoretical expectation that averaging predictions from diverse base models reduces the variance component of predictive error, at the cost of a moderate reduction in peak accuracy relative to the individual classifiers.

The lower accuracy of the Ensemble Method (84.25%) compared to the individual classifiers (99.33%) requires explanation. Probability averaging introduces a smoothing effect: when both base models assign high confidence to the correct class, the ensemble inherits this confidence. However, when the base models disagree, the averaged probabilities are pulled toward the decision boundary, increasing uncertainty and occasionally producing misclassifications. In this dataset, where the individual models achieve near-perfect separation, the ensemble's averaging introduces unnecessary noise rather than correcting complementary errors, a well-known limitation of homogeneous ensemble combination (i.e., combining models from the same algorithmic family). From a practical standpoint, the Ensemble Method's higher Kappa (0.6849) and MCC (0.6852) relative to its accuracy reflect good class-adjusted discrimination despite the accuracy gap, and its consistently low standard deviations indicate reliable, stable performance across data splits.

Discussion

Interpretation of Model Performance in Relation to Study Objectives

The study sought to ascertain the accuracy and reliability of credit risk assessment with the application of machine learning models such as Random Forest, Extra Trees Classifier and a probability average ensemble method in Ghanaian microfinance institutions. The results provide strong support for this aim. The average accuracy of both classifiers was 99.33% with AUC-ROC of 0.9997 on each of the ten stratified folds. This was higher than the reported accuracy of traditional rule-based and manual assessments in Ghana (Decardi-Nelson et al., 2014; Nyanzu & Quaidoo, 2018). These statistics reveal the uniformity and impartiality with which machine learning-based methods can identify high-risk borrowers, a feat that cannot be structurally attained through face-to-face interviews and personal judgement in credit evaluation for Ghana's micro-lending sector (Omowole et al., 2024).

The secondary aim of methodological rigour was addressed through fold-level oversampling isolation, 10-fold stratified cross-validation and full hyperparameter reporting. Design choices matter. For example, Chawla et al. (2002) demonstrated an artificial performance improvement by oversampling the test data by adding synthetic instances of the minority class to the test partition. The reported metrics are true generalisations of unseen real-world observations and not data leakage artefacts, since the RandomOverSampler is restricted to training partitions only. The methodological transparency of the present study is different from some previous studies, which oversample prior to the train-test split, which systematically overestimates the effectiveness of the models (Noriega et al., 2023).

Comparative Performance of the Three Models

The fact that the Random Forest and the Extra Trees classifiers give the same results for all seven evaluation metrics deserves interpretation. Both ensemble-based tree methods use the same randomisation principle for the features, but Extra Trees adds additional randomness in the split threshold selection process by sampling thresholds from the set of all features rather than looking for the best split (Geurts et al., 2006). Interestingly, this structural difference did not lead to differences in performance on the Tepa Man dataset. In this case, the loan-to-income ratio and interest rate are the most important signals in the data for prediction. They are so strong and discriminative that the additional threshold randomisation of Extra Trees did not provide any marginal informational advantage over the best-split strategy of Random Forest. This is consistent with the theoretical predictions that both algorithms will converge to similar decision surfaces when the number of features that are highly predictive of the classes is small compared to the total number of features (Ranglani, 2024), as observed in the datasets used in this study for both split strategies.

The performance of the individual classifiers is almost perfect (99.33%), while the ensemble method has an accuracy of 84.25%. The result is surprising, as it contradicts the usual belief that ensemble combination always yields better prediction performance. The probability-averaging strategy proposed here works best when the component models have complementary error patterns, i.e., when one model is right when the other is wrong and vice versa (Shah et al., 2023). However, the predictions in this work are highly correlated, as RF and ETC belong to the same family of algorithms and are trained on the same dataset. If the two models have similar prediction errors on observations at the decision boundary, probability averaging will exacerbate the error. This will push the predictions closer to the decision boundary and lower the confidence. This is well known in the literature on homogeneous ensembles: the result is a smoothing effect that reduces peak accuracy (Akinjole et al., 2024). In addition to the large proportionate reduction in accuracy, there is also a much greater stability indicated by the lower standard deviation across the folds (AUC = 0.0050) for the ensemble method (0.0061 SD across folds for the ensemble method compared with 0.0083 for the individual classifiers). This benefit of stability may outweigh the accuracy trade-off when a firm needs to deploy institutionally and requires consistent and predictable performance across a wide variety of borrower cohorts.

These comparative findings are in agreement with the body of general literature. The results of comparison of the tree-based ensemble models with the logistic regression and rule-based approaches for the loan default prediction reiterate the findings of Maindola et al. (2024) and Ranjan et al. (2024). Given the reported manual assessment baselines for Ghanaian micro-lenders, the models of the former significantly outperform the models of the latter. Romero and Sahoo (2024) also demonstrate that for financial applications, ensemble learning can simultaneously decrease both bias and variance, but this result is most relevant for heterogeneous ensembles comprising different algorithm families. Similarly, ensemble methods can enhance generalisation across domains depending on the diversity of base models (Alserhani & Aljared, 2023). The present study introduces a context-specific qualification: that in ensembles of individual high-performing models of the same family, stability and calibration metrics are more useful for determining the practical utility of the ensemble than accuracy metrics.

Interpretation of Near-Perfect Performance and Overfitting Considerations

The near-perfect classification metrics reported for RF and ETC require critical examination in relation to overfitting risk. Three explanations are consistent with the observed results and should be considered together rather than in isolation. First, the dataset may contain genuinely highly discriminative features that create clear class separation. The exploratory analysis confirms that loan-to-income ratio and interest rate are strongly predictive of default, and the feature importance rankings in Table 4 show that these two features alone account for approximately 47% of total Gini importance in both models. In financial datasets where default is strongly determined by a small set of structural features, near-perfect within-sample discrimination can be a legitimate reflection of true predictive signal rather than overfitting (Liu, 2024; Shukla, 2024). Second, the cross-validation design with fold-level oversampling isolation ensures that no synthetic data from the minority class contaminates any test fold, substantially reducing the most common source of data leakage in imbalanced classification tasks. Third, the consistency of performance across all ten folds, with standard deviations of 0.0083 for accuracy and 0.0009 for AUC, suggests that the metrics are not driven by a single unusually favourable data partition but reflect stable model behaviour across diverse subsets of the dataset.

Nevertheless, the single-institution origin of the dataset introduces an external validity constraint that accuracy and AUC metrics alone cannot resolve. The dataset is drawn exclusively from Tepa Man Microfinance Institution, and its borrower population may share socioeconomic and behavioural characteristics that create class boundaries more distinct than would be encountered across Ghana's broader microfinance sector. This concern aligns with findings from Fejza et al. (2022), who note that credit risk models trained on single-institution data in developing economies frequently overstate their generalisability because institutional-level factors, such as loan officer practices, product design, and target community characteristics, confound the learned decision boundary. Duman et al. (2023) similarly caution that performance metrics from single-source credit datasets require external validation before deployment claims can be substantiated. Accordingly, the near-perfect metrics reported here should be interpreted as establishing an upper bound on model capacity within the training distribution, not as a guarantee of equivalent performance on data drawn from other Ghanaian micro-lending institutions with different operational contexts.

Feature Importance and Alignment with Financial Distress Theory

The feature importance results are not only predictive but also theoretically significant. Financial distress theory suggests that the likelihood of default rises when debt service requirements reach a borrower's capacity to repay (Singh, 2024), and the loan-to-income ratio is directly predicted by it, since it is a variable of considerable importance in both models (around 28% of the Gini importance). The scatter plot analysis (Figure 1) finds that those who allocate over 30% of their income towards the loan repayment have a significantly higher tendency towards default in both the formal and informal credit markets, indicating that similar structural risk factors are present in both markets in Ghana. The second-place ranking of interest rate (around 19% importance) also underscores financial distress theory: higher

interest rates lead to lower residual income, which is available to repay the loan, and, as a result, compound the impact of high loan-to-income ratios, making such borrowers extremely risk-prone, as identified in the exploratory analysis (Kiralioğlu, 2024).

Like loan intent, the moderate importance of visibility in the distributional patterns of Figure 2 is theoretically based. The reason why debt consolidation borrowers have higher default rates could be that they may have been struggling financially before taking out the loan, while education loan and medical loan borrowers may have higher repayment motivation because of the returns of their loans or the need for medical care (Adesanya, 2024). This interpretation aligns with the study by Noriega et al. (2023), who show that in credit scoring models from several jurisdictions, loan purpose is a meaningful stratification variable. The dataset contained only a secondary source of discriminatory power, which could be a consequence of the generally similar collateral constraints of borrowers at Tepa Man Microfinance Institution, where formal loan collateral in the form of property ownership is uncommon. The limited predictive value of employment length (around 3% importance in both models) contrasts with its importance in some of the credit scoring models used in the West but may be due to the short and informal employment histories that are common among the borrowers in Ghana's micro-lending system (Oppong-Boakye et al., 2012).

The major disadvantage of using Gini importance when ranking features for this dataset is the notable overestimation of the importance for continuous and high-cardinality features, since there are more possible split points than with other features, giving them more opportunities to reduce impurity without any true predictive power (Nallakaruppan et al., 2024). Loan-to-income ratio and loan amount are both continuous and therefore, could be partially due to this methodological artifact as opposed to a true predictive signal. Permutation importance and SHAP values would generate more solid and model-agnostic estimates of importance and help clearly distinguish between the predictive contribution and the measurement bias in the importance scores in future work (Seitkulov et al., 2024).

Implications for Credit Risk Practice in Ghanaian Micro-Lending

The advantages of the accuracy and consistency of RF and ETC compared to manual evaluation are particularly clear in situations where loan officers are under time pressure to process a large number of applications and are likely to rely on intuition rather than evidence in their decision-making (Omowole et al., 2024). The models can be combined to minimise false positives (declined creditworthy borrowers that negatively affect portfolio growth and financial inclusion) and false negatives (defaults that threaten institutions' solvency) rather than replace the decision-making process. Ghanaian micro-lenders are challenged by high default and interest rates, leading to financial exclusion (Singh, 2024; Kiralioğlu, 2024).

By identifying loan-to-income ratio thresholds as important risk discriminants, this provides a practical, tangible, and workable basis for formalising underwriting standards. If the institution does not have the technical resources to deploy the full ML pipeline these rules based on the feature importance threshold and exploratory analysis can be used as temporary risk stratification criteria to capture a large part of the discrimination benefit without deploying the full ML pipeline. This finding is consistent with the proposed most affordable pathways for technological transfer in resource-constrained microfinance markets in Sub-Saharan Africa (Nwaimo et al., 2024; Gao et al., 2023).

Some structural features appear to have a substantial impact on the model's predictions in the context of regulation, raising some issues of transparency and fairness for borrowers. This may not be a reflection of borrower behaviour, but how the institution prices its loans, with the loan-to-income ratio and the interest rate together influencing the decision weight that makes up almost half of the model. Similar interpretability concerns have been raised in the credit scoring context, e.g., by Temelkov et al. (2024) who argued that ML systems, although accurate, may be exposed to regulatory risk due to lack of transparency in decision making. Such explanations are possible at the instance level (Seitkulov et al.,

2024; Nallakaruppan et al., 2024). They can assist lenders in assessing what types of loans are given to clients to make clearer financial decisions. Lenders can also educate such customers who want transparency in their financial transactions.

Limitations and Directions for Future Research

Several limitations of this study merit explicit acknowledgment. First, the dataset originates from a single institution, Tepa Man Microfinance Institution, and the generalisability of the trained models to other Ghanaian micro-lenders cannot be established without external validation on held-out institutional data. Second, no hyperparameter optimisation was performed; scikit-learn default settings were retained to enable controlled comparison of baseline model capacity, but marginal performance gains through grid search or Bayesian optimisation remain unexplored. Third, the homogeneous composition of the Ensemble Method, combining two models from the same algorithmic family, limits the error-correction potential of the ensemble strategy. Fourth, feature importance was computed exclusively using Gini impurity, which carries the high-cardinality bias limitation discussed above. Fifth, the class imbalance ratio of approximately 3.6:1, while addressed through RandomOverSampler, is relatively modest compared to real-world default rates at many microfinance institutions, and the models' behaviour under more severe imbalance conditions (e.g., 10:1 or 20:1) remains untested.

Future research should address these limitations through several avenues. Multi-institutional datasets spanning diverse Ghanaian micro-lending contexts would enable more robust assessments of model generalisability across borrower populations. Heterogeneous ensemble strategies combining RF or ETC with gradient boosting, neural networks, or logistic regression would introduce the model diversity necessary for ensemble combination to realise its theoretical performance advantages (Romero & Sahoo, 2024). The integration of alternative data sources, including mobile money transaction records, utility payment histories, and digital footprint data, could substantially improve credit discrimination for borrowers without formal financial records, an especially important direction given Ghana's rapidly growing mobile money ecosystem (Gao & Lau, 2024; Musyoka et al., 2024; Nwaimo et al., 2024). Longitudinal studies tracking model performance across macroeconomic cycles, including periods of inflation, currency depreciation, and agricultural output shocks relevant to rural Ghanaian borrowers, would evaluate whether the learned decision boundaries remain stable under distributional shift. Finally, the incorporation of fairness metrics, such as demographic parity and equalised odds across gender and geographic subgroups, would ensure that high aggregate accuracy does not mask disparate impact on protected borrower categories (Nallakaruppan et al., 2024).

Conclusion

This study investigated the application of Random Forest, Extra Tree Classifier, and a probability-averaged Ensemble Method to credit risk assessment in Ghanaian micro-lending institutions, using 32,581 loan records from Tepa Man Microfinance Institution. The study addressed methodological rigour by applying oversampling exclusively to training partitions, employing 10-fold stratified cross-validation, reporting full hyperparameter configurations and random seeds, conducting feature importance analysis, and critically examining near-perfect performance metrics in relation to potential overfitting and dataset characteristics.

The findings confirm that the loan-to-income ratio and interest rate are the primary predictors of default risk, with borrowers allocating more than 30% of income to debt service or facing interest rates above 10% exhibiting substantially elevated default probabilities. Loan intent emerged as a meaningful stratification variable, with debt consolidation and venture loans carrying higher risk profiles than education and medical loans. Homeownership type provided secondary discriminatory information, with renters showing greater default propensity relative to mortgage holders.

Among the ML models evaluated, both Random Forest and Extra Trees Classifiers achieved a near-identical mean accuracy of 99.33% and AUC-ROC of 0.9997, demonstrating exceptional within-sample discrimination on this dataset. The Ensemble Method, at 84.25% accuracy and AUC of 0.9231, exhibited stronger generalisability characteristics and lower performance variability across folds. These results collectively support the conclusion that ML models can substantially improve the accuracy, consistency, and reliability of credit risk assessment in Ghanaian micro-lending institutions, reducing dependence on subjective human evaluation.

Recommendations

Based on the study findings, the following recommendations are proposed for practitioners, policymakers, and researchers:

For micro-lending institutions: Machine learning models, particularly Random Forest and Extra Trees Classifiers, should be piloted as decision-support tools within existing credit evaluation workflows. Loan-to-income ratio thresholds should be formalised as underwriting criteria, and risk assessment protocols should be differentiated by loan intent, with stricter evaluation standards applied to debt consolidation and venture loan applications.

For model deployment: Prior to institutional deployment, models should be validated on fully held-out, external datasets from multiple micro-lending institutions to confirm out-of-sample generalisation. Ongoing performance monitoring is essential to detect distributional shifts in borrower populations over time.

For explainability: Future implementations should incorporate SHAP or permutation importance analyses to provide model-agnostic, instance-level explanations for individual lending decisions, supporting both loan officer understanding and regulatory compliance.

For policymakers: National financial regulatory bodies should develop guidelines for the responsible adoption of ML-based credit scoring in microfinance, including standards for model documentation, fairness assessment, and borrower transparency.

For future research: Comparative studies should investigate heterogeneous ensemble strategies combining RF or ETC with gradient boosting or neural network models to explore whether greater model diversity produces superior ensemble performance. Longitudinal studies tracking model performance across macroeconomic cycles are also warranted. The integration of alternative data sources, including mobile money transaction histories and utility payment records, offers a particularly promising direction for improving credit risk modelling for unbanked and underbanked populations in sub-Saharan Africa.

Acknowledgement

The authors would like to acknowledge all individuals whose constructive comments and suggestions helped to improve the quality of this paper. No external funding was received for the completion of this work.

Author Contribution Statement

Conceptualisation: PCA (lead), NBK (equal)

Data curation: SKA (lead), ROA (equal), CNO (equal)

Formal analysis: PCA (lead), ROA (equal), CNO (equal)

Methodology: SKA (equal), JD (equal)

Writing – original draft: PCA (lead), SKA (equal), NBK (supporting)

Writing – review & editing: PCA (equal), ROA (equal), JD (equal), NBK (equal)

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AI Disclosure Statement

No AI tool was used in this manuscript preparation.

References

Adesanya, M. E. (2024). Assessing credit risk through borrower analysis to minimize default risks in banking sectors effectively. *International Journal of Research Publication and Reviews*, *5*(12), 5479–5493. https://doi.org/10.55248/gengpi.5.1224.0233

Akinjole, A., Shobayo, O., Popoola, J., Okoyeigbo, O., & Ogunleye, B. (2024). Ensemble-based machine learning algorithm for loan default risk prediction. *Mathematics*, *12*(21), 3423. https://doi.org/10.3390/math12213423

Alserhani, F., & Aljared, A. (2023). Evaluating ensemble learning mechanisms for predicting advanced cyber-attacks. *Applied Sciences*, *13*(24), 13310. https://doi.org/10.3390/app132413310

Bin, J., Gardiner, B., Li, E., & Liu, Z. (2021). Extreme randomized trees for real estate appraisal with housing and crime data. In *Artificial Intelligence: Models, Algorithms and Applications* (pp. 83–97). Bentham Science Publishers. https://doi.org/10.2174/9781681088266121010008

Chafale, J. H., Wadetwar, R. S., Giriya, D. M., Badhiye, S., Borkar, P., & Shinde, S. (2024). Credit risk analysis using machine learning. *2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 1–5. https://doi.org/10.1109/ICCUBEA61740.2024.10774950

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Decardi-Nelson, I., Asamoah, O., Solomon-Ayeh, B., & Nduro, K. (2014). The informal sector and mortgage financing in Ghana. *Ghana Journal of Development Studies*, *9*(2), 136. https://doi.org/10.4314/gjds.v9i2.8

Duman, E., Aktas, M. S., & Yahsi, E. (2023). Credit risk prediction based on psychometric data. *Computers*, *12*(12), 248. https://doi.org/10.3390/computers12120248

Fejza, D., Nace, D., & Kulla, O. (2022). The credit risk problem: A developing country case study. *Risks*, *10*(8), 146. https://doi.org/10.3390/risks10080146

Gao, T., & Lau, R. Y. K. (2024). Leveraging deep learning and multimodal signals from social media to enhance credit risk prediction. *2024 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 1–6.

Gao, X., Yang, X., & Zhao, Y. (2023). Rural micro-credit model design and credit risk assessment via improved LSTM algorithm. *PeerJ Computer Science*, *9*, e1588. https://doi.org/10.7717/peerj-cs.1588

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Jayaram, E. S. (2024). Machine learning-based loan default prediction: Models, insights, and performance evaluation in peer-to-peer lending platforms. *Educational Administration: Theory and Practice*, 12975–12989.

Kiralıoğlu, T. (2024). Investigating the use of machine learning in automating credit scoring for microfinance. *Human Computer Interaction*, *8*(1), 29.

Kofi Oppong-Boakye, P., Ahinful, G. S., & Danquah, J. B. (2012). Micro-loans and micro enterprises in Ghana: Players, roles and challenges. *European Journal of Business and Management*, *4*(20).

Kumar, S. (2024). Enhanced fraud detection in financial transactions using hyperparameter-tuned random forests. *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–7.

Kyriazos, T., & Poga, M. (2024). Application of machine learning models in social sciences: Managing

nonlinear relationships. *Encyclopedia*, *4*(4), 1790–1805.

Li, C., & Zhang, J. (2024). Research on credit risk prediction models based on machine learning. *2024 6th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, 1–4.

Liu, S. (2024). Application analysis of machine learning models in credit risk. *Highlights in Science, Engineering and Technology*, *107*, 76–81.

Maindola, M., Al-Fatlawy, R. R., Kumar, R., Boob, N. S., Sreeja, S. P., Sirisha, N., & Srivastava, A. (2024). Utilizing random forests for high-accuracy classification in medical diagnostics. *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, 1679–1685.

Musyoka, G., Waititu, A., & Imboga, H. (2024). Credit risk modelling using RNN-LSTM hybrid model for digital financial institutions. *International Journal of Statistical Distributions and Applications*, *10*(2), 16–24.

Nallakaruppan, M. K., Chaturvedi, H., Grover, V., Balusamy, B., Jaraut, P., Bahadur, J., Meena, V. P., & Hameed, I. A. (2024). Credit risk assessment and financial decision support using explainable artificial intelligence. *Risks*, *12*(10), 164.

Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. *Data*, *8*(11), 169.

Nwaimo, C. S., Adegbola, A. E., & Adegbola, M. D. (2024). Predictive analytics for financial inclusion: Using machine learning to improve credit access for underbanked populations. *Computer Science & IT Research Journal*, *5*(6), 1358–1373.

Nyanzu, F., & Quaidoo, M. (2018). *Access to finance constraint and SMEs functioning in Ghana* (MPRA Paper No. 83202). Munich Personal RePEc Archive.

Omowole, B. M., Urefe, O., Mokogwu, C., & Ewim, S. E. (2024). Strategic approaches to enhancing credit risk management in microfinance institutions. *International Journal of Frontline Research in Multidisciplinary Studies*, *4*(1), 053–062.

Ranglani, H. (2024). Empirical analysis of the bias-variance trade off across machine learning models. *Machine Learning and Applications: An International Journal*, *11*(4), 01–12.

Rani, R., & Gupta, S. (2024). Predicting home loan approvals using random forest classifiers. *2024 3rd International Conference for Advancement in Technology (ICONAT)*, 1–4.

Ranjan, A., Gourisaria, M. K., Singh, J. P., Panda, A. R., Mishra, S. R., & Parihar, V. (2024). Leveraging advanced machine learning algorithms for loan repayment prediction. *2024 15th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–6.

Romero, B., & Sahoo, D. (2024). Integrated learning comprehensive evaluation of stock market prediction. *Journal of Research in Science and Engineering*, *6*(12), 39–40.

Saha, T., Biswas, S. K., Sanyal, S., Verma, N., & Purkayastha, B. (2023). Credit risk prediction using extra tree ensembling technique with genetic algorithm. *2023 14th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–5.

Seitkulov, Y., Sickory Daisy, J., Deepshika, S., & G., H. (2024). Credit risk analysis using explainable artificial intelligence. *Journal of Soft Computing Paradigm*, *6*(3), 272–283.

Shah, M., Gandhi, K., Patel, K. A., Kantawala, H., Patel, R., & Kothari, A. (2023). Theoretical evaluation of ensemble machine learning techniques. *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 829–837.

Shukla, D. (2024). A survey of machine learning algorithms in credit risk assessment. *Journal of Electrical Systems*, *20*(3), 6290–6297.

Singh, K. (2024). The dual nature of peculiar problems in microfinancing: Perspectives on market efficiency and public policy nexus. *Journal of Economic and Administrative Sciences*. Advance online publication.

Su, Y. (2024). Financial technology credit risk modeling and prediction based on random forest algorithm. *2024 Second International Conference on Data Science and Information System (ICDSIS)*, 1–4.

Syam, N., & Kaul, R. (2021). Random forest, bagging, and boosting of decision trees. In *Machine*

Learning and Artificial Intelligence in Marketing and Sales (pp. 139–182). Emerald Publishing Limited.

Temelkov, Z., & Georieva Svrtinov, V. (2024). AI impact on traditional credit scoring models. *Journal of Economics*, *9*(1), 1–9.

Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.

Wang, Q., Nguyen, T.-T., Huang, J. Z., & Nguyen, T. T. (2018). An efficient random forests algorithm for high-dimensional data classification. *Advances in Data Analysis and Classification*, *12*(4), 953–972.

Yan, G. (2024). Research on the application of alternative data in credit risk management. *Highlights in Business, Economics and Management*, *40*, 1156–1160.

Zhou, Z.-H. (2021). Computational learning theory. In *Machine Learning* (pp. 287–313). Springer Singapore.