Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 8 (2026) 2649

Journal of the Nigerian Society of Physical Sciences

Integrated data-driven credit default prediction in Uganda using machine learning models

George Muddu*, Shefiu Olusegun Ganiyu, Adekunle Olugbenga Ejidokun, Yusuf Abass Aleshinloye

Department of Computer Science, Kampala International University, Uganda

Abstract

The prediction of credit facility defaulters is quite a challenge in Uganda, particularly for those without a formal banking history. Existing prediction models cater for prediction using conventional banking records which is not sufficient. The use of integrated data to cater for the unbanked population is required to further enhance financial inclusivity and stability in Uganda's financial landscape. This study therefore aims at filling this gap by using machine learning techniques on a rich blend of financial data, including mobile money and Fintech (Financial Technology) services, as well as traditional banking records. Several machine learning algorithms used for loan default prediction were compared, such as Random Forest, Logistic Regression, Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost). Random Forest Model showed 96.66% accuracy, 79.65% recall, 96.52% precision and 0.85 AUC. XGBoost model was found to have an accuracy of 95.23%; recall, 73.32%; precision, 94.11%; and Area Under the Curve (AUC) of 0.81. However, Random Forest performed best by all metrics with XGBoost following slightly. Logistic Regression showed 89.53% accuracy but had a very low recall at 43.24% and precision at 66.59%. SVM performed averagely with 93.21% accuracy and 62.80% recall all falling below that of XGBoost and Random Forest. Thus, the study revealed the significance of machine learning models like Random Forest and XGBoost for credit scoring prediction. Overall, it will improve the ability of institutions and policymakers to identify potential default borrowers so as to mitigate loan default rates and ensure economic growth in underserved communities through more accurate and inclusive credit evaluation tools.

DOI:10.46481/jnsps.2026.2649

Keywords: SVM, XGBoost, Credit defaulter, Prediction model, Financial data

Article History :
Received: 29 January 2025
Received in revised form: 13 July 2025
Accepted for publication: 14 July 2025
Available online: 08 October 2025

Communicated by: Oluwatobi Akande

1. Introduction

Financial sector of Uganda demonstrates quite a remarkable growth in digital financial services with the coming in of various mobile money platforms like MTN and Airtel [1]. Such services have tremendously expanded their frontiers, enabling majority to transact and access a financial system. Without emphasizing much on the utility of mobile money, for example, the aforementioned researches, [2] shed light on how the emerging mobile money ecosystem has transformed financial transactions in the country as a basis of adopting data-driven credit scoring systems. The study reveals that data based on mobile money transactions can improve not only the accuracy of credit scoring models but also their accessibility in understanding borrowers' creditworthiness with more depth. Despite the progress made,

*Corresponding author Tel. No.: +234-803-801-7995.

Email address: mlawgeorge@gmail.com (George Muddu)

it is evident that there are missing mechanisms to assess the creditworthiness of the individuals particularly those who do not have any credit history through traditional banking [3].

Traditionally, creditworthiness in Uganda has been assessed using methods like loan officers' subjective judgments and the use of basic historical records. For instance, Zofi cash uses employee's salary and workplace while other fintech uses alternative credit assessments, such as bank statements and mobile money statements [4]. In addition, others use the Credit Reference Bureau (CRB) services. The intent behind the creation of CRB was to increase the performance of loans by allowing the lenders to have access to greater market data in their credit evaluation choices [5]. These methods often failed to capture the full financial behaviour of individuals, especially those in the informal sector who rely heavily on mobile money or other Fintech services [6]. Such traditional methods can be proned to bias and inaccuracy, leading to both unjust denial of credit and overextension to unreliable borrowers. Otherwise, credit scoring models using machine learning could actually provide a better answer for the aforementioned challenges.

More specifically, it has been shown by the Bank of Uganda's statistics that default rates on microloans stood at 16%, which would later rise to over 22% by the year 2021 [7]. Bank of Uganda statistics also indicated default rate increment by 14.6% year-on-year on a net basis on household loans [8]. Such behavior creates distrust for the financial system and discourages responsible borrowers from fulfilling their repayment obligations. Such behavior discourages lenders from offering loans to new customers. Hence, much has not been achieved concerning assessing creditworthiness. Borrowers have, therefore, suffered high-interest rates for all loans on account of the inability to sort quite well between reliable borrowers and high-risk individuals [9]. This has led to a negative impact on overall economic growth, as potential borrowers, especially from unbanked segments, have been excluded from the formal credit system due to robust credit scoring systems.

Machine learning proves to be much more efficient over the traditional credit assessment techniques. Merging common sources of information like mobile money transactions, Fintech accreditations, and traditional bank records, machine learning imparts an entirely new view of the credit profile. Such models can discover interesting patterns in financial behavior that human loan officers or simple rule-based systems might miss, thereby creating a more comprehensive understanding of the risk associated with individual borrowers in determining more individualized interest rates. This improves credit assessment accuracy while reducing default risk, helping to bring down risk premiums on loans and consequently achieve lower interest rates for prudent borrowers [10]. Healthy financial ecosystems reward responsible borrowing and flag potentially troublesome borrowing behavior.

Furthermore, using machine learning is very pertinent for the Ugandan largely unbanked populations who are mostly reliant on informal systems as well. However, the challenges to the adoption of machine learning technology in Uganda are many. For examples, insufficient financial data, lack of robust regulatory framework and the adaptation of machine learning models to local practices in financial transactions, are key areas to be considered. The research work aims to use integrated data for predicting loan defaults in Uganda that has a largely unbanked population and evaluate the impact of the data on various performance metrics across different ML algorithms. Resolving the gaps in these areas will build a strong case for the application of integrated data in credit scoring and encourage the advancement of financial inclusion in Uganda.

2. Related Work

He *et al*. [11] focused on personal loan default rates with comparison using two machine learning models Random Forest (RF) and LightGBM. The dataset used was obtained from Datafountain's official website hosted by the China Computer Federation and Central Plains bank was used containing 750,000 records. LightGBM outperformed RF with an AUC of 86% as compared to RF's 55%. LightGBM is one of the best algorithms for predicting loan defaults.

Zhu *et al*. [12] focuses on loan default prediction while addressing the black-box issues to enhance interpretability and transparency. The dataset was provided by Datawhale and obtained from the website https://tianchi.aliyun.com/competition/entrance/531830/information. The dataset uses 612,742 sample and applied Logistic Regression, Decision Tree (DT), XGBoost, and LightGBM models. Local Interpretable Model-Agnostic Explanations (LIME) were used for interpretability. LightGBM achieve the highest AUC and precision which provided insights into key variables influencing loan defaults.

Kozina He *et al*. [13] focused on prediction of leasing contract defaults from leasing companies using 4500 cases analyzed using RF, AdaBoost, Gradient Boosting and Deep Neural Networks. The models were evaluated using precision and recall with Deep Neural networks achieving excellent recall and random forest with the best precision for non-defaults. The model requires more input variables to enhance the performance of the metrics used.

A comparative studies of machine learning algorithms such as logistic regression, Decision Tree, Random Forest and Gradient Boosting Machine was conducted using a bank dataset was used for survey consisting of 10000 credit accounts [14]. Comparison shows GBM attains the best AUC and accuracy amidst all the algorithm while a similar comparison done with a different bank data shows SVM and RF acquires 100% precision and accuracy shows 98.34% and 98.2% using the PCA dimensionality reduction technique [15]. All including Suhadonik *et al*. [16] reported significant improvements in default prediction accuracy when compared to traditional credit scoring methods.

Naveen *et al*. [17] focuses on credit prediction in Small and Medium-sized Enterprises (SMEs) using Random Forest (RF), AdaBoost, Gradient Boosting, XGBoost, and Linear Discriminant Analysis (LDA). The dataset used was obtained from Centre for Monitoring Indian Economy Pvt Ltd. (CMIE) dataset for SMES. RF achieved an accuracy of 92.19%, recall of 92.42% and AUC-ROC of 100% showing a superior prediction over other algorithms.

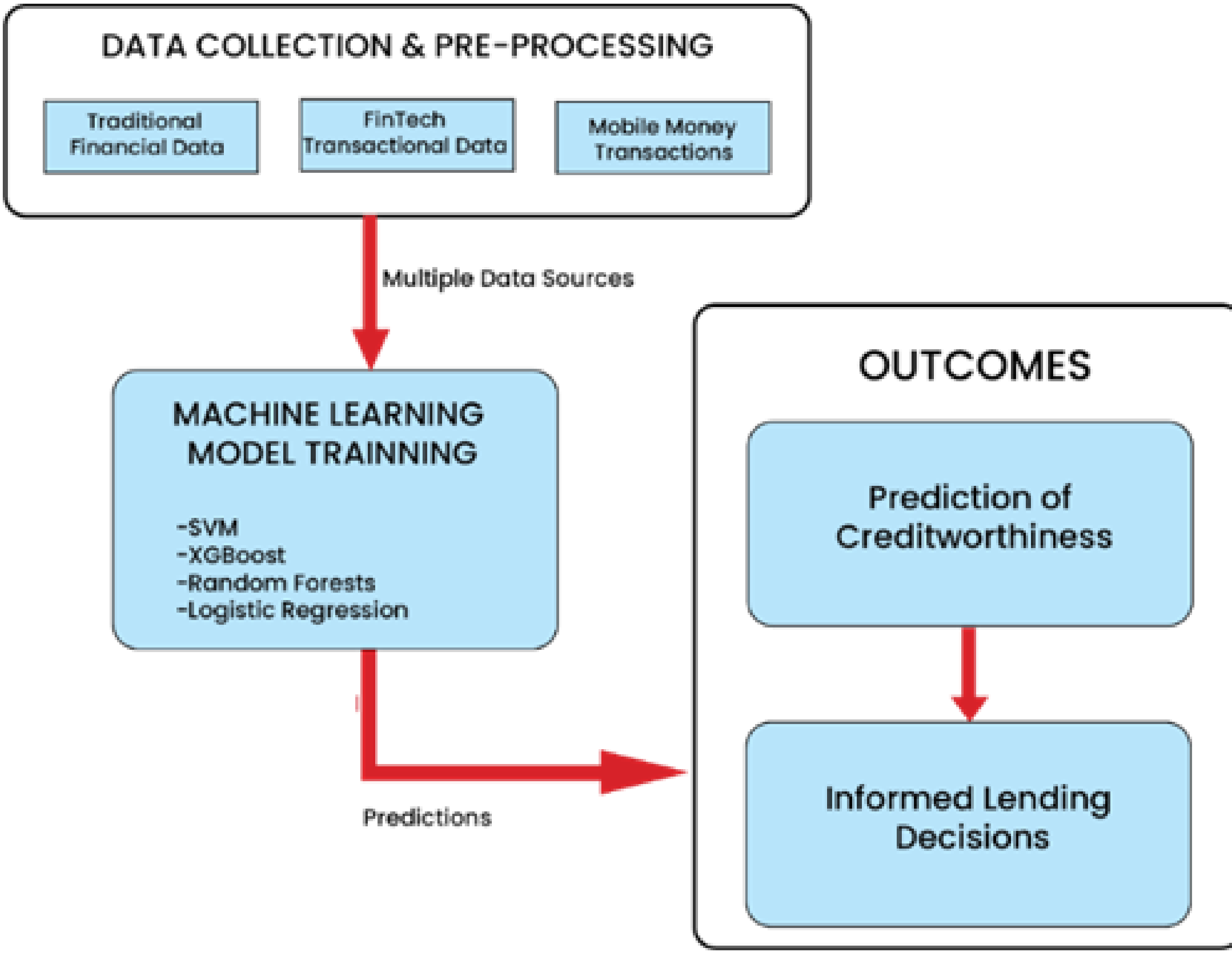

Figure 1. Conceptual framework for enhancing credit scoring through machine learning.

[18] predicted loan default in Chinese P2P market using RF, Extreme Gradient Boosting Tree (XGBT), Gradient Boosting Model (GBM), and Neural Network (NN) on 54,477 loans dataset obtained from Renrendai.com. Synthetic Minority Oversampling Technique (SMOTE) was used to balance the imbalanced dataset. Accuracy for all the algorithms was above 90% and both literatures acknowledged RF having superior prediction over other algorithms. [19] also proved RF better than DT.

Most research on loan default prediction focuses on datasets from China, India and other developed or emerging economies. However, Machine learning applications in credit scoring are not yet researched in recent studies from emerging markets such as Uganda. Most of the existing credit scoring models rely on traditional data about the customers. Nevertheless, using alternative/integrated financial data such as such as mobile money transactions, telecom data can enhance the model's accuracy and partaking in the provision of surplus information in credit and lending in credit scoring models [21]. Hence, this study.

3. Methodology

3.1. Conceptual framework

Figure 1 depicts the conceptual framework that illustrates the process by which machine learning enhances credit scoring systems, particularly in contexts like Uganda where traditional credit data is limited. Each component of the diagram represents a crucial part of the overall framework.

3.2. Data collection

The primary data sources include transaction and loan repayment records from a Bike loan provider (which deals in bike financing), X-Fintech company (which specializes in financing various products through loan), blinded data from two telecommunications companies in Uganda (to gather information on mobile money transactions), banking records from a bank in Uganda. Table 1 shows the number of records and date range for data utilized in this research.

Table 1. Data quantities and time frames.

Provider	Number of Record	Date Range
Bike Loan Provider	30,000 loan repayment records	January 2020 - December 2023
X-Fintech Company	30,000 loan repayment records	January 2020 - December 2023
Telecom A and Telecom B	100,000 mobile money and 100,000 loan repayment records	January 2020 - December 2023
Bank X	50,000 banking and 50,000 loan repayment records	January 2020 - December 2023

Table 2. Telecom transactions.

S/N	**Attributes**
1	Transaction ID
2	Timestamp
3	Amount
4	Sender Phone Number
5	Receiver Phone Number
6	Transaction Type

Table 3. FinTech transactional data.

S/N	Attributes
1	User ID
2	Loan Amount
3	Repayment Status
4	Loan Duration
5	Interest Rate
6	Payment History

Table 4. Traditional banking records.

S/N	**Attributes**
1	Account Number
2	Transaction Date
3	Transaction Amount
4	Account Balance
5	Transaction Type (e.g., debit, credit),
6	Branch Code

Table 5. Demographic data.

S/N	Attributes
1	Name
2	Date of Birth
3	Gender
4	National ID
5	Address
6	Phone Number

Table 6. Boda boda loan provider.

S/N	Attributes
1	Customer ID
2	Transaction Date
3	Transaction Type
4	Amount
5	Financial Institution
6	Customer Name
7	Phone Number
8	Account Number
9	Account Type
10	Principal Amount
11	Interest Rate
12	Loan ID
13	Repayment Schedule
14	Fees

The dataset used was anonymized and name of organizations are also withheld to ensure confidentiality and privacy of proprietary data in accordance with relevant data protection regulations.

3.3. Data structure for primary data and data integration

Table 2-6 shows the dataset from different sources having different structures and patterns. The table shows the raw data description of the structure each dataset before fusion.

The process of preparing and refining data for analysis involved several critical steps, each designed to ensure the dataset was robust, consistent, and optimized for modelling. Initially, data integration was achieved by creating a unique customer identifier, which served as a cornerstone for unifying disparate datasets. This identifier was constructed by combining common fields, such as phone numbers and account numbers, which were present across multiple datasets. However, this task was not without challenges, as inconsistencies in data formats and missing identifiers posed significant hurdles. To overcome these, standardization techniques were employed to unify date formats, numeric representations, and categorical labels, ensuring consistency across the datasets. Additionally, fuzzy matching techniques were utilized to resolve discrepancies in identifiers, such as partial or incomplete phone numbers, allowing for more accurate data linkage.

Another key aspect of the data preparation process was addressing missing data, which is a common issue in large datasets. Missing values were handled thoughtfully to maintain the integrity of the data. For numerical data, statistical methods such as imputing the mean or median were used to fill in gaps, ensuring that the dataset remained representative. In cases where missing values were deemed irrelevant, they were dropped entirely to avoid introducing noise. For categorical data, missing values were either replaced with the mode—the most frequent category—or encoded as a separate category to preserve the dataset's structure and avoid bias in subsequent analyses.

To enhance the dataset's predictive power, feature engineering was undertaken to uncover hidden patterns and improve model performance. Several new features were generated to capture meaningful insights from the raw data. For instance, transaction frequency was calculated as the count of transactions per user over a specific period, providing a measure of user activity. The average transaction amount was computed as the mean value of a user's transactions, offering insight into their financial behaviour. Additionally, a loan repayment ratio was derived to represent the proportion of loans repaid on time, serving as an indicator of creditworthiness. A financial stability score was also created, based on trends in account balances, to reflect a user's financial health over time. These engineered features enriched the dataset, enabling more nuanced and accurate modelling.

Finally, data alignment was performed to consolidate information from various sources into a cohesive dataset. This was achieved by merging the data based on the unique customer identifier, ensuring that records from different sources were accurately linked. During this process, certain columns, such as bio data—including name, date of birth, and gender—were deemed irrelevant for modelling purposes and were excluded from the final dataset. This step streamlined the data, focusing only on features that were directly relevant to the analytical objectives. Through these combined efforts—data integration, handling missing data, feature engineering, and data alignment—the dataset was transformed into a robust and well-structured resource, ready for advanced analysis and modeling.

3.4. Data cleaning and pre-processing

Data pre-processing is a key phase in ensuring data quality. Given the diverse nature of the datasets, ranging from mobile money transactions to loan repayment records, the data cleaning process begins with handling missing values, duplicates, and outliers, which are common in financial datasets, especially those collected from various sources. Missing values were addressed using the pandas library for data manipulation. Mean and median replacement were done to ensure imputed values aligned with the data distribution. Some records that included too many missing values were also dropped. Duplicate entries were identified and removed. Outliers were also removed or treated to minimize their influence on the model's accuracy. Feature encoding was done using one hot encoder to convert categorical variables into numerical values for variables with multiple categories to ensure that the model could effectively interpret these features. Features of varying scales were normalized and standardized. The Standard Scaler function standardized the dataset for a mean of zero and a standard deviation of one-an important step for gradient boosting algorithms, which tend to be sensitive to the scales of the features. The datasets were split into training and testing sets by the 'train_test_split function of scikit-learn, through which 80% of the data was used for training the model and 20% of the data was held back for testing. This function allowed it to perform a stratified split so that the distribution of the target variable was seen across training and testing sets. Performance metrics of the model for accuracy and precision thus reflect the generalization capabilities of the model to unseen data. The SMOTE function was also set for the appropriate parameters so that they could adjust the level at which oversampling was done. Thus, SMOTE also evaded the unfair, complex issues created by imbalanced data so that the credit scoring model developed in its wake does become fairer and stronger with more accuracy in terms of all classes of credit defaults.

3.5. Feature Engineering

After data cleaning, the subsequent task is to combine and align the data into an integrated dataset. Table 7 describes the dataset with detailed transactional records for various financial activities.

It was observed that each company had loan and transaction profiles that could be found in other data sets before the data was blinded. Cross-referencing was done to create a whole, comprehensive and interconnected dataset. It started with X-Fintech Company and Bike Loan Provider Y, then used the loan profiles in these two Fintechs to inform the profiles collected from Telecom A, Telecom B, and Bank X. This ensured a rather unwavering integration of a heterogeneous financial behaviour across several platforms. The crucial step toward an exhaustive and cohesive dataset for analysis is data integration from various sources. Due to privacy concerns and data consistency concerns, those columns that include any personal attribute like name, date of birth, gender, address, and national ID number are being dropped. Besides, this has no contribution in terms of predictions since it does not affect model performance. Jobs such as account managers, loan officers, etc. which relate to organizational workflows, and do not have any impact on scoring model have also been dropped.

Mostly, the generated features through engineering have been ingested into the dataset. The Total Deposits and Withdrawals feature represented the total amount of money that had been deposited or withdrawn from fintech entities. The Total Repayments and Disbursements feature accounted for the overall disbursement and repayment amounts associated with the various companies.

Frequency of withdrawals and deposits has indicated how frequently individuals make transactions regarding their accounts in terms of deposits or withdrawals. Similarly, the Frequency of Repayments and Disbursements illustrated how frequently a customer had made repayments towards loans or received disbursements. Lastly, the default Indicator was a categorical variable that had indicated the assessed credit worthiness of the customer, this was achieved through K-means clustering.

Table 8 provides an overview of attributes (or features) that describe a customer's financial activities with various fintech entities. The features are used in training the model.

3.6. Machine learning model development for credit scoring

This investigation was primarily focused on Gradient Boosting Machines (GBM), with special reference to the XGBoost algorithm, including other algorithms like logistic regression, decision trees, and support vector machines for the initial

Table 7. Raw data description.

S/N	Attributes	Description
1	Customer ID	A unique identifier for each customer
2	Transaction Date	The date on which the transaction occurred.
3	Transaction Type	The type of transaction for example deposits, withdrawals, repayments, or disbursements
4	Amount	The monetary value of the transaction.
5	Financial Institution	The financial institution associated with the transaction. For example, Telecom A, Telecom B, Bank X, Bike Loan Provider Y, and X-Fintech Company
6	Customer Name	Name of the customer
7	Phone Number	Contact number
8	Account Number	Unique number for the customer's account
9	Account Type	Type of account (e.g., savings, checking, mobile money).
10	Principal Amount	The original amount of the loan.
11	Interest Rate	Interest rate applied to the loan
12	Loan ID	Identifier for the loan associated with repayments or disbursements.
13	Repayment Schedule	Schedule for loan repayments.
14	Fees	Any fees or charges associated with the transaction.

Table 8. Feature engineered datasets.

S/N	**Attributes**	**Description**
1	Total Deposits and Withdrawals	Total amount of money deposited or withdrawn from Fintech Entity like Telcom A, Telecom B, Bank X, X-Fintech Company, and Bike Loan Provider Y
2	Total Repayments and Disbursements	Total amount of money disbursed or repaid to Telcom A, Telecom B, Bank X, X-Fintech Company, and Bike Loan Provider Y
3	Frequency of Withdrawals and Deposits	How often a customer performs deposits or withdrawals transactions with their accounts
4	Frequency of repayments and disbursements	How often a customer makes repayments towards loans or receives disbursements
5	Default Indicator	A categorical variable indicating the credit worthiness of the customer

phase to set a baseline against the others. The objective function, which forms the core of the XGBoost, comprises two important parts: loss function and regularization term. The loss function measures the degree of deviation from the predicted to the actual values, while the regularization term controls model complexity to avoid overfitting. The additive model has each subsequently added tree taking successional steps at minimizing the objective function to make predictions better. The regularization would also impose certain penalties for overly complicated models and hence another compromise between accuracy and simplicity.

Data is split at nodes in order to create the trees and thus satisfies the objective function. A best split is found on the basis of maximized "gain," which signifies a reduction in the objective function with the help of the first and second derivatives of the loss function.

The mathematical formulation of such an objective function for XGBoost would be given in said equation (1).

i. Objective function: The objective function in XGBoost consists of a loss function and a regularization term. The loss function measures the difference between the predicted and actual values, while the regularization term controls the complexity of the model to avoid overfitting.

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \Omega(f_k), \tag{1}$$

where l is the loss function (e.g., mean squared error for regression, logistic loss for classification), $\hat{y}_i^{(t)}$ is the prediction for the i-th instance at the t-th iteration and $\Omega(f_k)$ is the regularization term for the k-th tree.

ii. Additive model: As stated in equation (2) XGBoost builds the model in an additive manner. Each new tree f_t is added to minimize the objective function.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i). \tag{2}$$

iii. Regularization term: The regularization term Ω in equation (3) helps to control the complexity of the trees.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \tag{3}$$

where T denote the number of leaves in the tree, w_j is the weight of the j-th leaf and γ and λ are regularization parameters.

iv. Tree structure: At every node, the split is done to minimize the objective function as shown in equation (4) to build the tree structure. The maximum gain, which is the reduction in the objective function, finds out a split optimal.

$$\text{Gain}=\frac{1}{2}\left[\frac{(\sum_{i\in L} \; g_i)^2}{\sum_{i\in L} \; h_i+\lambda}+\frac{(\sum_{i\in R} \; g_i)^2}{\sum_{i\in R} \; h_i+\lambda}-\frac{(\sum_{i\in L\cup R} \; g_i)^2}{\sum_{i\in L\cup R} \; h_i+\lambda}\right]-\gamma \tag{4}$$

where g_i is the first derivatives of the loss function with respect to the prediction $\hat{y}_i$, h_i is the second derivatives of the loss function with respect to the prediction $\hat{y}_i$, L represent the left branches after the split and R represent the right branches after the split.

This structure gives XGBoost a lot of leverage to improve prediction accuracy in controlling model complexity, thereby proving itself to be perfect for credit scoring in the Ugandan financial sector. In short, this systematic evaluation of all the potential algorithms resulted in the choice of XGBoost as the central contender because of its successful reputation in credit scoring applications. The result was to have irrefutably chosen a model by matching it against all essential metrics according to how well it performed on the selection process. By this way, the chosen model was also in the line of providing solutions to the study objective of improving the credit decision process in Uganda's financial sector.

3.7. Model training

Machine learning models were implemented in Jupyter Notebook and Python a endowed with 32 GB RAM, Intel Core i7-1355U Processor at clock speed 3.4 GHz, and 1 TB disk space. Windows 11 was employed as the OS for experimentation.

Several machine learning models were trained using Logistic Regression, Random Forest Classifier, LightGBM, Support Vector Machine (SVM), and Neural Network. Logistic Regression served as a baseline model using a one-vs-rest (OvR) approach for multi-class classification, trained on standardized data and evaluated on the test set.

4. Result and discussions

Initial exploratory data analysis revealed key characteristics of the dataset, customers usually make 125 to 200 withdraws per year from Telecom A, Telecom B, Bank X and customers make average repayments of 700,000 UGX to 1,000,000 repayments to X-Fintech Company. Transaction volumes varied significantly across different financial institutions, with Telecom A and Telecom B processing the highest number of transactions.

4.1. Deposits and withdrawals

Customer deposits and withdrawals across different fintech entities exhibit significant variation. With Telecom B, total deposits per person ranged from nothing to nearly 9 million UGX, averaging around 2.5 million UGX. Total withdrawals per person varied from 0 to over 5 million UGX, with an average of about 1.5 million UGX. For Telecom B, total deposits per person reached up to 14 million UGX, averaging around 3.5 million UGX, while total withdrawals averaged around 2.5 million UGX but could go as high as 9.5 million UGX. At the commercial bank, total deposits per person ranged from 0 to 7.4 million UGX, and total withdrawals varied from 0 to 5.1 million UGX. Average deposits for commercial bank were about 2 million UGX, while withdrawals averaged around 1.2 million UGX.

As part of the Exploratory Data Analysis (EDA) for developing a credit score prediction model, the distribution of various financial activities across different platforms (Telecom A, Telecom B, Bank X, Bike Loan Provider Y, X-Fintech Company) was examined. The histograms presented provided a clear overview of these distributions.

The histograms below in Figure 2a to Figure 2d represent the distribution of the different financial activities and transactions distribution of various financial activities across different platforms was examined.

Each histogram showed a right-skewed distribution, indicating that the majority of customers engaged in low levels of financial activities such as deposits, withdrawals, disbursements, and repayments, while a smaller number of customers handled significantly higher amounts. This pattern was consistent across all platforms and for both the total amounts and the frequency of transactions. This consistent right-skewed distribution suggested similar customer behaviour patterns across different financial service providers, which would be crucial for identifying key features and trends in building the credit score prediction model.

4.2. The correlation matrix

Figure 3 reveals several insights into the relationships between different financial variables. High positive correlations were evident among related transactions, such as between total deposits and total disbursements for each service provider. For instance, total Telecom B deposits showed a strong positive correlation with total Telecom B disbursements and total Telecom B repayments. Similarly, high correlations were observed between total Telecom A deposits and total Telecom A disbursements, as well as between total Bank X deposits and total Bank X disbursements.

The "Credit Assessment" column showed notable correlations with several variables. It had a positive correlation with net deposits across various service providers, indicating that higher net deposits were associated with a better credit assessment. There were also moderate correlations between the credit assessment and repayment frequencies, suggesting that more frequent repayments were linked to a better credit score.

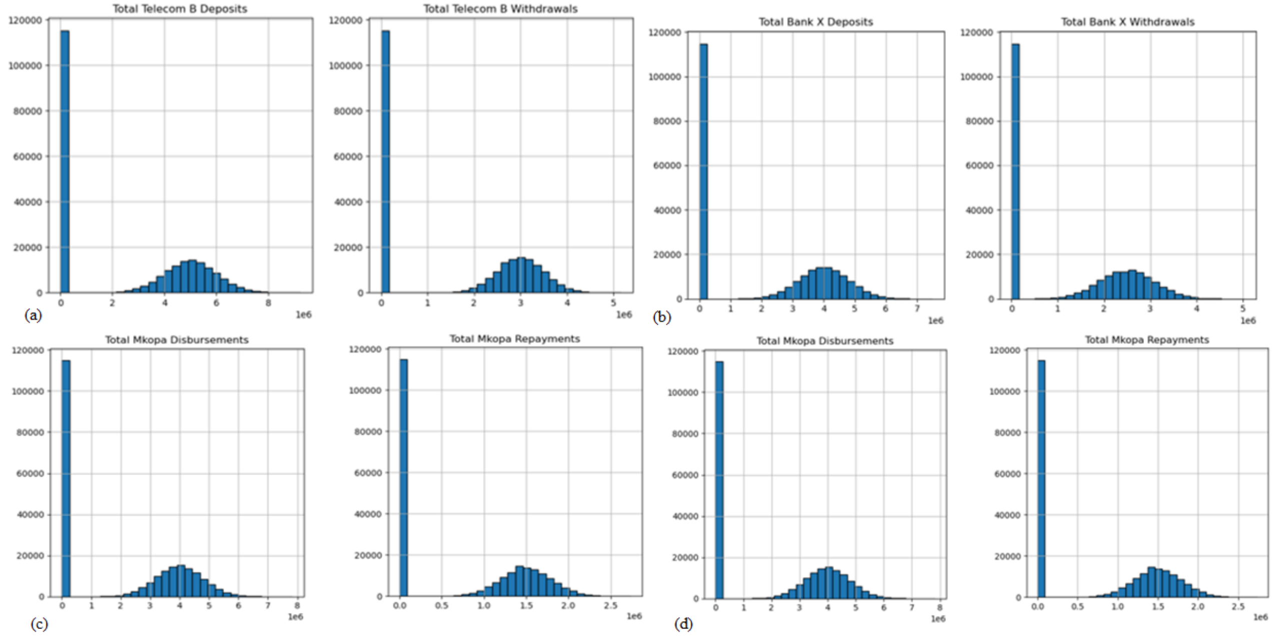

Figure 2. (a) Distribution of Telecom B deposits and withdrawal (b) Distribution of Total Bank X deposits and withdrawal (c) Distribution of Total Mkopa disbursements and repayments (d) Distribution of Telecom A deposits and withdrawal.

Interestingly, the heatmap displayed negative correlations in some areas. For example, withdrawal frequencies for different service providers had a negative correlation with deposit frequencies and repayment frequencies, indicating that higher withdrawal activities might reduce deposit and repayment activities.

4.3. *Model evaluation for machine learning algorithm selection*

The Table 9 and Figure 4 presents a comparative analysis of various machine learning models based on their performance metrics for a classification task. Six models—Logistic Regression, Random Forest, LightGBM, SVM, XGBoost, and a Neural Networks.

The Random Forest model achieved the highest accuracy (96.66%), recall (79.65%), and precision (96.52%), indicating that it was the most effective at correctly classifying both defaulters and non-defaulters. LightGBM closely followed, with similar metrics, demonstrating its strength as a high-performing gradient boosting model. The Neural Network also performed well, particularly in precision (87.98%) and recall (77.63%), showcasing its ability to generalize complex patterns in the data.

In contrast, the Logistic Regression model, while achieving a decent accuracy (89.53%), had lower recall (43.24%) and precision (66.59%), suggesting that it struggled to identify all defaulters. The SVM model and XGBoost provided a good balance between accuracy (93.21%) and AUC (84.94%), but its recall was relatively lower (62.80%).

The Neural Network demonstrates strong performance as illustrated in Figure 4a with an Area Under the Curve (AUC) of 0.81 for Class 0 ('low default'), 0.89 for Class 1 ('medium default'), and 0.86 for Class 2 ('high default'). These AUC values indicate that the Neural Network is particularly effective at distinguishing 'medium default' cases, with a high AUC of 0.89, suggesting a strong ability to predict these scenarios accurately. While the model performs well in predicting 'high default' cases, it is slightly less effective with 'low default' cases, as reflected in the lower AUC for Class 0.

The LightGBM model exhibits similar AUC values to the Neural Network, with 0.81 for Class 0, 0.89 for Class 1, and 0.86 for Class 2. This model's performance closely mirrors that of the Neural Network as it can be visibly seen or compared in Figure 4a and Figure 4e, especially in predicting 'medium default' and 'high default' scenarios. The consistency in AUC values between LightGBM and the Neural Network suggests that both models are well-suited to your credit prediction task, particularly in identifying higher-default cases.

On the other hand, the Logistic Regression model shows slightly lower performance compared to the other models as illustrated in Figure 4c. It has an AUC of 0.78 for Class 0, 0.84 for Class 1, and 0.80 for Class 2. While still a viable model, Logistic Regression is less effective at distinguishing between the different default classes, particularly with 'low default' and 'high default' cases. According to the AUC values identified for this model, it is likely to face a little more difficulty in making a correct prediction compared to the more sophisticated models such as LightGBM and the Neural Network.

The Support Vector Machine model performs really well

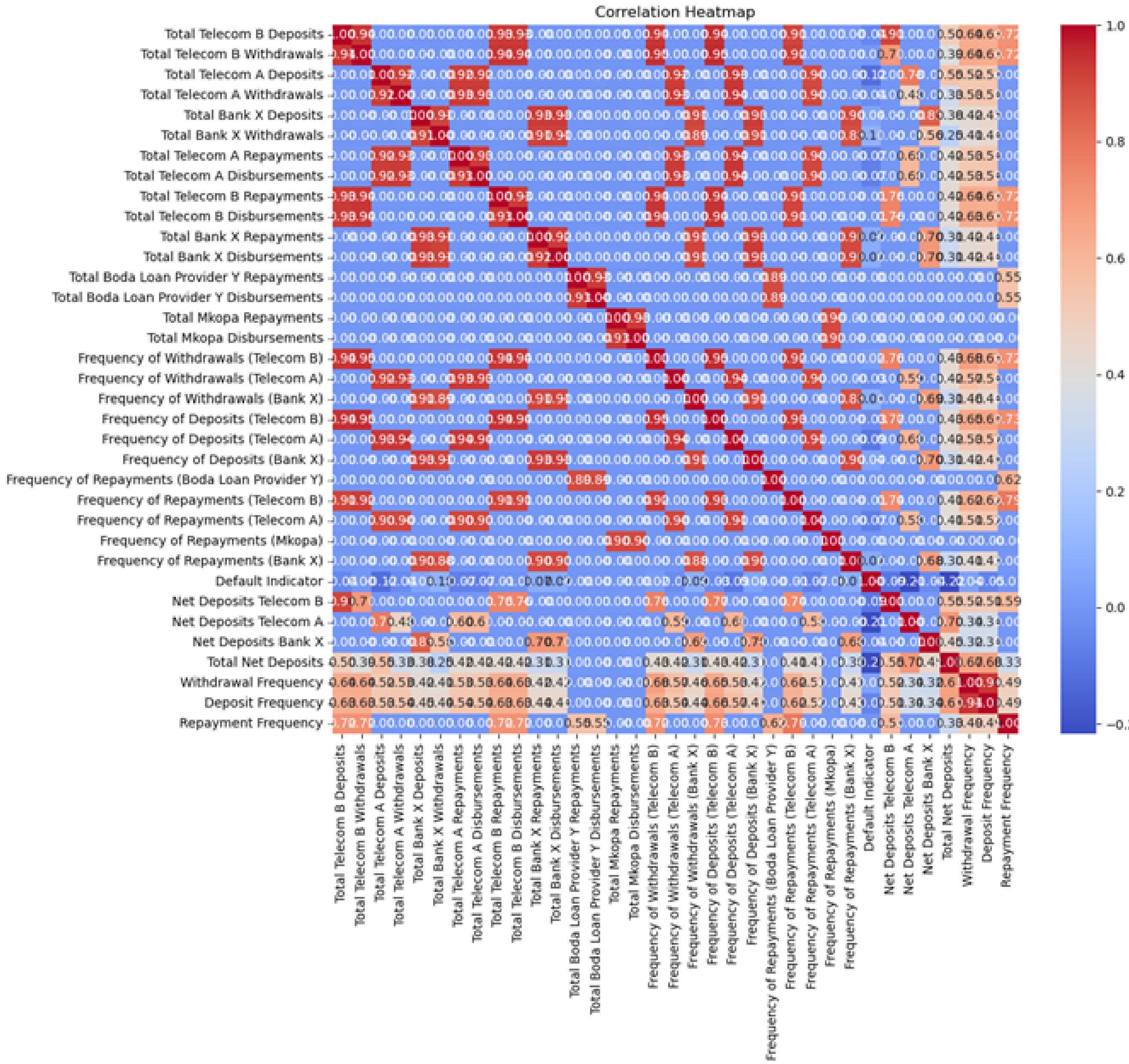

Figure 3. Correlation matrix.

Table 9. Model performance comparison

Model	Accuracy	Recall	Precision	AUC
Logistic Regression	0.895261	0.432447	0.665914	0.808249
Random Forest	0.96658	0.796489	0.965172	0.853467
LightGBM	0.965913	0.793169	0.962783	0.852824
SVM	0.93213	0.628002	0.856754	0.849375
XGBoost Model	0.952334	0.733233	0.941107	0.811934
Neural Network	0.953536	0.776312	0.879849	0.853495

with an AUC of 0.81 for Class 0, 0.88 for Class 1, and 0.86 for Class 2 as shown in 4f. SVM turns out to be another strong performer for 'medium default' cases, producing an impressive AUC value of 0.88. The model falls slightly behind the Neural Network and LightGBM in catching the targets of 'low default' and 'high default' cases but, for sure, emerges competitive enough as a choice for your credit prediction model.

The Random Forest model has also shown quite good per-

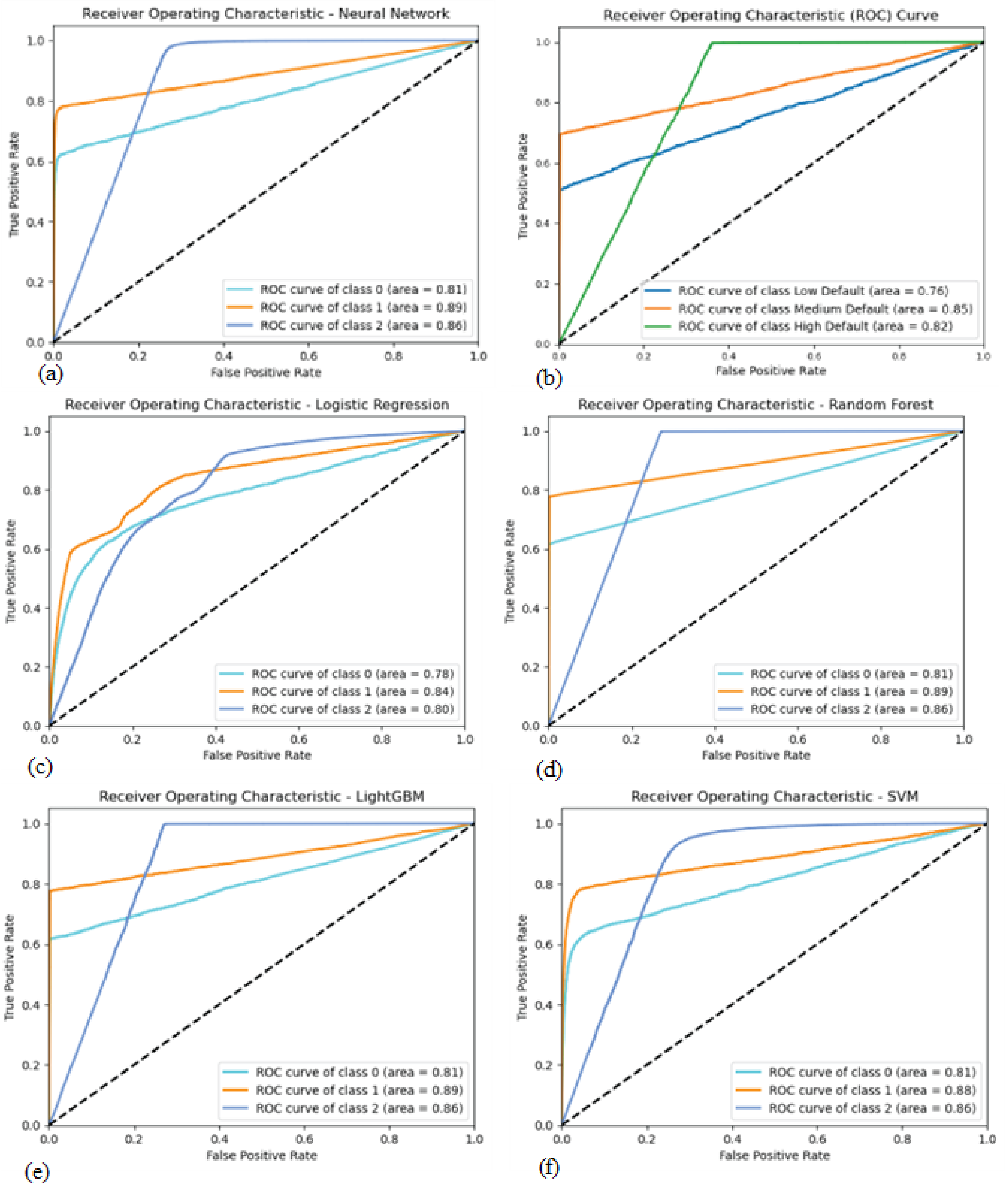

Figure 4. (a) ROC - Network (b) ROC - XGBoost (c) ROC - Logistic Regression (d) ROC - Random Forest (e) ROC - LightGBM (f) ROC - SVM.

formance, with AUC values of 0.81 for Class 0, 0.89 for Class 1, and 0.86 for Class 2. Along with the other top performing models, Random Forest was able to achieve great success in predicting 'medium default' and 'high default'. Given its bal-

Table 10. Model comparison

Studies	Accuracy	Recall	Precision	AUC
In this Study (XGBoost)	95.23%	73.32%	94.11%	0.81
In this Study (Random forest)	96.66%	79.65%	96.52%	0.85
Wang et al. (2020) (Random Forest)	95.29%	76.66%	95.29%	0.88
Wang et al. (2020) (Logistic Regression)	81.01%	13.84%	61.16%	0.56
Wang et al. (2020) (Naïve Bayes)	79.99%	0.09%	36.00%	0.50
Kumar et al. (2021) (XGBoost)	94.1%	82%	90%	0.97
Yang et al. (2019) (XGBoost)	89.4	78%	93%	0.92

anced performance among classes, the model can be trusted for the credit prediction task, particularly in extreme cases when a higher default is expected.

The XGBoost model also provides an attractive mix of models in Figure 4(b) with 95.23 percent accurate, a high precision of 94.11 percent, and a solid AUC score of 81.19 percent. The consistency of results in accuracy, precision, and AUC indicates the robustness of the model, especially when it comes to correctly classifying non-defaulters and this, without accuracy compromise. However, Random Forest performs best across all metrics. Although recall has a value of 73.32 percent which is slightly low for Random Forest, the high precision and lack of reliance on AUC are indicative of a model that gives minimum false positives and false negatives. Its overall balance of precision, recall, and accuracy makes a strong choice, able to rely on performing significantly across a number of scenarios. Thus, with that performance guarantee, Random Forest and XGBoost are the algorithms to use for credit prediction problems when looking for good performance across evaluation metrics.

The model currently being studied for RF and XGBoost shows very high performance on several parameters such as feature importance, AUC, precision, recall, and F1-score. The performance of this two models is also competitive with previous studies, especially with high-default predictions, in terms of precision and recall. Therefore, it can be understood that with high accuracy and reliable default categorization, the RF and XGBoost model is a good candidate for complex financial prediction tasks.

4.4. Model comparison

The Table 10 provides a comparative analysis of various machine learning models based on their performance metrics in classification tasks. The metrics used to evaluate these models include Accuracy, Recall, Precision, and AUC (Area Under the Curve), which offer a comprehensive view of each model's effectiveness.

Machine learning in credit scoring has opened up much promise and many challenges for nearly all emerging economies like Uganda. However, much more research needs to be conducted for machine learning techniques to be specifically aligned with the unique socio-economic and regulatory contexts of such contexts to gain full benefits from more advanced credit scoring models.

5. Conclusion

Machine learning provides an alternative to predict credit facility defaulters through flexible models based on credit seekers' financial history. Leveraging these machine learning models; credit lenders and borrowers have an objective mechanism to assist credit decision-making. Especially, the models generated from the XGBoost algorithm showed remarkable performances for predicting credit defaulters using datasets harmonised from a commercial bank, mobile money wallets, and Fintech companies. The application of such models will ensure that only credit-worthy loan applicants access credit facilities. This will no doubt increase the survival rates of loan providers and the overall well-being of the financial sector of Uganda. Moreover, the models provide an avenue for stakeholders in the financial sector to make data agonistics policies to monitor loan services in the country, rather than using subjective approaches which inhibit the growth of the sector. However, the dataset employed in this study was sourced from a few of the operators in the financial sector of Uganda. Thus, the findings are limited to the dataset providers and size. Also, the results are restricted to the machine learning algorithms experimented in the study.

Modern advanced ML models are superior in performance. Ensemble approach to ML has been proven to assist in effectively resolving issues regarding imbalanced datasets, timely detection and cost-effective approach [21, 22]. However, the black box brings about challenges since financial institutions need transparency of the credit decisions which is not available in complicated models leading to the need for tools like SHAP and LIME for transparently making model outputs somewhat understandable to the stakeholders [23].

Recently, studies have focused on entailing fairness metrics and mechanisms for detecting bias in credit scoring models; this would avoid fairness and bias problems within credit scoring processes and also grant results in a fairer manner in markets that are under-served like Uganda, where disadvantaged and marginalized ones usually face financial hurdles brought about by systemic challenges [15]. Another recent study which is under-researched is the use of tools that will enable financial institutions to easily plug into XAI methods to improve model transparency and trust among both regulators and consumers [24]. Scalability presents a challenge with complex models due to the computational cost generation, which necessitates optimization for practical application [25]. The emergent markets' concerns, such as lacking reliable data, are left out since most of their studies tend to be focused on the developed markets.

Therefore, future research could explore much wider comparative studies of ML algorithms for accuracy and efficiency, readability, and overall improvement toward adoption and larger datasets from other financial institutions and informal financial service providers.

Data availability

The data supporting the findings of this study are not publicly available for research purposes due to privacy and ethical restrictions outlined in agreements with the involved organizations. They may be accessed upon request from the corresponding author, subject to approval and compliance with these restrictions.

References

[1] M. G. Kasim & T. Matsumoto, "Mobile money, remittances, and household welfare: Panel evidence from rural Uganda", World Development **79** (2016) 127. https://doi.org/10.1016/j.worlddev.2015.11.006.

[2] T. Suri & W. Jack, "The long-run poverty and gender impacts of mobile money", Science **354** (2016) 1288. https://doi.org/10.1126/science.aah5309.

[3] A. Demirguc-Kunt, L. Klapper, D. Singer & S. Ansar, "The global findex database 2017: measuring financial inclusion and the fintech revolution", World Bank Publications **1** (2018) 126033. http://documents.worldbank.org/curated/en/332881525873182837.

[4] Zofi Cash, "Ugandan fintechs in measuring credit worthiness", 2023. [Online]. https://furtherafrica.com/2023/06/08/uganda-fintech-startup-zofi-cash-receives-us1m-pre-seed/.

[5] P. Mukiibi, P. Mugambe & J. Kampumure, "The effect of credit reference bureau information sharing on credit assessment in financial institutions in Uganda", Ugandan Journal of Management and Public Policy Studies **17** 39. https://ujmpps.umi.ac.ug/index.php/ujmpps/article/view/47.

[6] M. Andrianaivo & K. Kpodar, "ICT financial inclusion, and growth: evidence from African countries", 2011. [Online]. https://www.imf.org/external/pubs/ft/wp/2011/wp1173.pdf.

[7] A. Irumba & Z. Enock, 'Role of financial literacy in decreasing loan default rates among micro-borrowers in Uganda", Metropolitan Journal Of Academic Multidisciplinary Research **3** (2024) 200. https://www.researchgate.net/publication/384323510_Role_Of_Financial_Literacy_In_Decreasing_Loan_Default_Rates_Among_Micro-Borrowers_In_Uganda.

[8] Monitor, "Rein in growing loan default rate", 2023. [Online]. https://www.monitor.co.ug/uganda/oped/editorial/rein-in-growing-loan-default-rate-4397274.

[9] B. Ssekiziyivu, R. Mwesigwa, M. Joseph & I. Nkote Nabeta, "Credit allocation, risk management and loan portfolio performance of MFIs—A case of Ugandan firms", Cogent business & management **4** (2017) 1374921. http://dx.doi.org/10.1080/23311975.2017.1374921.

[10] E. Nzibonera & I. Waggumbulizi, "Loans and growth of small-scale enterprises in Uganda: a case study of Kampala Central business area", African Journal of Business Management **14** (2020) 159. http://dx.doi.org/10.5897/AJBM2020.8985.

[11] Y. He, Y. Jian, T. Liu & H. Xue, "Optimization of machine learning models for prediction of personal loan default rate", International Conference on Bigdata Blockchain and Economy Management **2022** (2022) 270. https://doi.org/10.2991/978-94-6463-030-5_29.

[12] X. Zhu, Q. Chu, X. Song, P. Hu & L. Peng, "Explainable prediction of loan default based on machine learning models" Data science and management **6** (2023) 123. https://doi.org/10.1016/j.dsm.2023.04.003.

[13] A. Kozina, Ł. Kuźmiński, M. Nadolny, K. Miałkowska, P. Tutak, J. Janus & R. Król, "The default of leasing contracts prediction using machine learning", Procedia Computer Science **225** (2023) 424. http://dx.doi.org/10.1016/j.procs.2023.10.027.

[14] Tianyi Xu, "Comparative analysis of machine learning algorithms for consumer credit risk assessment", International Conference on Computer Engineering, Information Science & Application Technology **4** (2024) 60. https://doi.org/10.62051/r1m3pg16.

[15] H. Suresh & J. Guttag, "Understanding potential sources of harm throughout the machine learning life cycle", MIT Case Studies in Social and Ethical Responsibilities of Computing [Preprint], 2021. https://doi.org/10.21428/2c646de5.c16a07bb.

[16] N. Suhadolnik, J. Ueyama & S. Da Silva, "Machine learning for enhanced credit risk assessment: An empirical approach", Journal of Risk and Financial Management **16** (2023) 496. https://doi.org/10.3390/jrfm16120496.

[17] K. Naveen, N. Pavitha & K. Ashutosh, "Machine learning for credit default prediction in SMES: a Study from emerging economy, Accountancy Business and the Public Interest **7** (2021) 54. https://abpi.uk/wp-content/uploads/2024/07/04SP2406.pdf.

[18] J. Xu, Z. Lu & Y. Xie, "Loan default prediction of Chinese P2P market: a machine learning methodology", Scientific Reports **11** (2021) 18759. https://doi.org/10.1038/s41598-021-98361-6.

[19] H. J. Bature, D. D. Wisdom, T. T. Dufuwa & I. O. Ayetuoma, "Credit default prediction system using machine learning", Computology: Journal of Applied Computer Science and Intelligent Technologies **3** (2023) 51. https://doi.org/10.17492/computology.v3i1.2305.

[20] R. Hlongwane, K. Ramaboa & W. Mongwe, "Enhancing credit scoring accuracy with a comprehensive evaluation of alternative data", **19** (2024) e0303566. https://doi.org/10.1371/journal.pone.0303566.

[21] E. G. Dada, A. I. Birma & A. A. Gora, "Ensemble machine learning algorithm for cost-effective and timely detection of diabetes in Maiduguri, Borno State", Journal of the Nigerian Society of Physical Sciences **6** (2024) 2175. https://journal.nsps.org.ng/index.php/jnsps/article/view/2175/345.

[22] E. Idongesit, U. L. Chinedu & A. E. Ibor, "A stacked ensemble approach with resampling techniques for highly effective fraud detection in imbalanced datasets", Journal of the Nigerian Society of Physical Sciences **7** (2025) 206. https://journal.nsps.org.ng/index.php/jnsps/article/view/2066/366.

[23] S. Lundberg & S. I. Lee, "A unified approach to interpreting model predictions", presented at Conference on Neural Information Processing Systems, California, U.S., 4th December, 2017. http://dx.doi.org/10.48550/arXiv.1705.07874.

[24] L. Noriega, M. Zhang & J. Fernandez, "Comparative analysis of machine learning algorithms in default prediction: Decision trees, random forests, and support vector machines", Journal of Credit Risk Analytics **34** (2023) 98. http://dx.doi.org/10.13140/RG.2.2.31652.14725.

[25] I. Steinwart & A. Christmann, "Support Vector Machines", in *Support Vector Machines*, Springer, Berlin, Germany, 2008, pp. 164–201. https://link.springer.com/book/10.1007/978-0-387-77242-4.