BENIN JOURNAL OF PHYSICAL SCIENCES
BJPS Vol. 2(1), June, pg. 159-174 (2025)
ISSN 3043-6931(Print)
3043-694X(Online)
www.bjps.org.ng

Predicting Likelihood for Loan Default Among Loan App Borrowers: A logit Classification Approach

B. E. OMOKARO and R. E. EMUDIAGA*

Department of Statistics, Delta State Polytechnic, Otefe, Oghara, Nigeria

Received: 12/04/2025 ***Accepted:*** 30/06/2025

Abstract

Predicting the likelihood of missed loan repayments is essential for banks and credit providers to effectively manage lending risks and promote responsible financial services. This research utilises a binary classification approach, specifically logistic regression, to categorise borrowers based on their potential to fail in meeting repayment obligations. The predictive framework includes a range of economic and personal attributes such as age, employment type, income level, loan size, repayment period, the number of current loans, applicable interest rate, and whether the borrower resides in a rural or urban setting. Data were obtained through a combination of loan repayment histories and structured questionnaires, covering a total of 397 individuals. Analytical procedures were executed in R programming. Key predictors influencing the probability of repayment failure include age, loan size, tenure, the count of existing credit facilities, and interest rate. Conversely, individuals with higher earnings were less likely to fall behind on repayments. The predictive tool demonstrated strong classification ability, achieving an overall correctness rate of 87.9%, a precision rate of 99.4%, a sensitivity of 88.1%, and an ROC, AUC value of 0.8029, suggesting solid differentiation between reliable and risky borrowers. As a result, the study advocates for more rigorous borrower vetting, loan structuring aligned with income capacity, and consideration of advanced machine learning models for refining credit risk analysis.

Keywords: *Loan default prediction, Logistic regression, Credit risk assessment, R programming, Borrower characteristics, Predictive modeling.*

*Corresponding author e-mail: emudiagaeric@gmail.com

1. INTRODUCTION

The digital lending industry has grown exponentially in Nigeria over the past decade, driven by increased internet penetration, financial technology (fintech) innovations, and the rising demand for quick and accessible credit. Conventional financial institutions, due to their stringent eligibility criteria and prolonged processing times, have created a void in access to credit, particularly affecting individuals with low income and owners of micro and small enterprises. This gap has paved the way for the widespread adoption of digital lending platforms, which provide quick, collateral, free loans with minimal paperwork. Despite their convenience and popularity, these platforms now face rising challenges linked to non, repayment, which could jeopardise their operational viability and long, term profitability. Identifying the key drivers of missed repayments is, therefore, essential for digital lenders aiming to manage exposure to credit risk and promote sustainable lending practices.

Loan default occurs when borrowers fail to meet their repayment obligations within the stipulated time. In the case of digital lending, where loans are often disbursed without physical collateral, defaults pose significant financial risks to lender (Awuza et al., 2022). Additionally, the privacy of online lending makes it challenging for lenders to assess borrowers' creditworthiness effectively, further exacerbating the default problem. Although certain financial technology providers utilise advanced tools such as artificial intelligence and data, driven insights to assess loan applicants, there is still an evident absence of uniform and locally adapted credit evaluation frameworks suitable for Nigeria's unique lending environment.

The regulatory landscape for online lending in Nigeria remains underdeveloped. While agencies like the Central Bank of Nigeria (CBN) and the Federal Competition and Consumer Protection Commission (FCCPC) have introduced measures to regulate digital lenders, enforcement remains a challenge. Many loan apps operate without proper licensing, charging exorbitant interest rates and using aggressive debt recovery tactics. The absence of a well, structured credit reporting system further complicates risk assessment, as there is limited data on borrowers' credit history, making it difficult to identify repeat defaulters.

In response to these challenges, predictive modeling techniques, such as logistic regression analysis, have been increasingly used in financial risk assessment. Logistic regression helps identify significant predictors of loan default and estimates the probability of a borrower defaulting based on their financial and demographic characteristics (Nureni and Adekola, 2022). By applying statistical models to primary data collected directly from borrowers, this study seeks to build an evidence, based framework for assessing online loan default risk among Nigerians.

The rapid growth of online lending platforms in Nigeria has provided individuals and small businesses with easy access to credit. However, the increasing rate of loan defaults has raised concerns among lenders, financial institutions, and regulators. Many borrowers fail to meet repayment obligations due to factors such as income instability, financial mismanagement, high interest rates, and fraud. These defaults threaten the sustainability of digital lending, leading to abusive calls, borrower's defamation of character, stricter lending conditions, higher interest rates, and financial losses for lending institutions.

Despite advancements in financial technology, predicting loan default risk remains a challenge due to the lack of reliable credit history among many borrowers, inconsistent economic conditions, and limited access to verifiable financial data. Traditional credit scoring methods often fail to capture the diverse financial behaviors of Nigerians using online loan apps.

Mantey et al. (2021) carried out an empirical investigation aimed at identifying the main variables influencing personal loan default rates, using borrower data obtained from a community, based bank in Ghana. The study analyzed responses from 196 loan recipients, focusing on attributes such as family size, academic qualifications, employment type, gender, age, and marital condition. Utilizing the Cox Proportional Hazards model, they evaluated how each factor contributed to the chances of missed loan repayments. Their findings revealed that academic background, gender, age, and marital status had no statistically significant influence on repayment behavior. However, the number of dependents and type of employer were shown to be crucial. Specifically, each additional dependent increased the borrower's default probability by 21.025%, signaling the economic burden of supporting larger households. Moreover, individuals employed outside the public sector were found to be 84.118% more prone to default, emphasizing the protective role of stable, government, backed employment in ensuring repayment reliability. These insights reinforce the need to incorporate personal and occupational data when assessing borrowers' creditworthiness in rural banking systems.

In another study, Efekodo et al. (2025) explored how cutting, edge machine learning algorithms can enhance the forecasting of loan repayment defaults beyond the capabilities of traditional methods. Working with a dataset of 50,000 loan applicant records, which included a variety of personal, financial, and loan, related variables, they partitioned the data into a 70:30 training and testing split for model development. They evaluated several predictive models, including Decision Trees, Random Forests, Gradient Boosting, Logistic Regression, and

Gaussian Naive Bayes. Of all the models tested, GaussianNB demonstrated superior accuracy at 78.8% on the test dataset. It proved particularly effective at capturing intricate data relationships and minimizing errors in classification. The authors recommended greater adoption of such algorithms to improve risk analysis in credit operations, while also advocating for the use of alternative data inputs—such as customer behavior patterns and broader economic indicators—as well as deep learning techniques in subsequent studies to further boost predictive performance.

Similarly, Nureni and Adekola (2022) focused on enhancing banks' ability to anticipate loan defaults as a strategy to reduce the occurrence of Non, Performing Assets (NPAs) and optimize profitability. Their work compared multiple analytical techniques to determine the most robust method for predicting borrower delinquency. Leveraging publicly available loan data from Kaggle, they trained and validated several machine learning models, assessing each using performance metrics like Precision, Accuracy, Recall, and F1, Score. The analysis involved eight algorithms, but the logistic Regression emerged as the top performer, with accuracy rates of 83.24% and 78.13% across two different datasets. Naïve Bayes followed closely with 82.16% and 77.34%, respectively. These results demonstrate the substantial utility of both traditional and machine learning approaches in predicting loan default likelihood, underlining the importance of model selection in enhancing the precision of credit risk management frameworks used by financial institutions.

Anam et al. (2024) carried out an extensive review of existing studies on how machine learning algorithms (MLAs) have been applied in recent years, specifically from 2020 to 2023, to improve the prediction of loan defaults. Their review traced the evolution from conventional statistical models toward more sophisticated ML techniques within the credit risk evaluation landscape. The analysis underscores the banking industry's increasing dependence on MLAs for improving prediction reliability. Among the algorithms assessed, Random Forest was consistently recognized for its robustness in handling intricate data and delivering high accuracy. Kaggle was identified as a primary data source, reinforcing the importance of accessible, well, structured datasets in building dependable models. The review also recommended future explorations in the areas of big data integration, the application of ensemble learning strategies, and the deployment of deep learning frameworks. Despite limitations such as the study's focus on a specific time window and reliance on available datasets, the authors stressed the need for continuous exploration to keep up with innovations in loan default prediction and credit risk modeling.

In another study, Adewole et al. (2016) examined environmental influences on loan defaults, aiming to identify and rank contributing factors, assess default trends across various industries, and suggest improved credit management practices. Surveying 120 loan recipients spanning 13 economic sectors, they applied logistic regression to measure the impact of ten key environmental factors on default likelihood. Their findings revealed sector, specific disparities, with the service industry recording the highest rate of loan defaults. Complementing their survey, they also analyzed data from the Central Bank of Nigeria's 2015 defaulters' list. The study concluded that banks should adopt more cautious and tailored lending strategies for sectors with heightened default risks. This research highlights the value of sector, based risk assessments in curbing credit losses and promoting financial stability.

Similarly, Eweoya et al. (2019) focused on detecting fraudulent activities in loan administration through machine learning. Utilizing the Naïve Bayes classifier, their study aimed to overcome the shortcomings of traditional fraud detection systems, which are often inefficient and prone to errors. Unlike earlier works that examined either loan eligibility or fraud in isolation, this study incorporated fraud detection into the broader context of loan default prediction. Trained on labeled data, the Naïve Bayes model reached an accuracy of 78%, proving its effectiveness in spotting fraudulent patterns linked to loan defaults. The authors advocated for the broader use of ML tools in financial institutions to bolster fraud mitigation strategies and improve overall risk assessment practices.

Addressing the broader issue of increasing default rates, Akinmoluwa et al. (2024) investigated the application of multiple machine learning models to better screen loan applicants. Using a dataset sourced from Zindi Africa, they evaluated four models, Decision Tree, Gradient Boosting Classifier, Random Forest, and Gaussian Naïve Bayes, based on performance metrics such as Accuracy, Precision, Recall, F1, Score, and the Confusion Matrix. Among these, the Gaussian NB model demonstrated the best performance, achieving an accuracy rate of 77%. These results emphasize the growing utility of machine learning in improving loan application vetting, reducing default probability, and empowering banks with data, driven decision, making tools for credit approval processes.

Jahanzaib et al. (2024) undertook an in, depth evaluation of influential publications on credit risk forecasting frameworks released from 2015 to 2024. Their analysis covered diverse modeling techniques such as text, based systems, integrative review papers, blended algorithms, intelligent prediction tools, and conventional statistical methods. The research highlights the growing importance of addressing credit repayment failures, which pose a serious challenge to global financial resilience. Through a structured and data, driven assessment, the study maps out current practices, identifies methodological shortcomings, and tracks the development of predictive techniques over time. In addition to summarizing leading innovations in the domain, the review outlines strategic areas for further investigation, offering fresh perspectives to refine credit evaluation practices. Notably, the work distinguishes itself by concentrating on well, recognized and frequently cited academic sources from renowned journals, ensuring its credibility and impact. The insights derived carry practical value for industry professionals, financial regulators, and scholars committed to enhancing institutional safeguards against future credit defaults.

Awuza et al. (2022) explored how factoring in early loan repayment behavior can improve the effectiveness of models used to forecast loan default risks. Although many banks now use machine learning tools to evaluate borrowers, these tools often fail to consider the predictive value of early repayment habits. The study emphasizes that integrating this variable can lead to more accurate and reliable predictions. To test this, six different supervised learning methods, Random Forest, Artificial Neural Network, Classification and Regression Tree, Support Vector Machine, Logistic Regression, and Naïve Bayes, were deployed to build comparative models. Each algorithm was applied to two sets of data: one with the early repayment feature and one without. The models' effectiveness was gauged using multiple indicators, including accuracy, precision, recall, Root Mean Square Error (RMSE), and the Receiver Operating Characteristic (ROC) score. Findings showed that models which included early repayment data consistently outperformed their counterparts, with the Random Forest algorithm achieving the highest performance, scoring 93% in accuracy, 90% in precision, 89% in recall, 11% RMSE, and an ROC score of 81%. This suggests that acknowledging repayment timing adds valuable insight to credit risk evaluations.

Despite extensive research on loan default prediction, significant gaps remain in understanding the unique factors influencing online loan app defaults in Nigeria. Many studies have focused on traditional bank loans (Mantey et al., 2021) or datasets from global repositories like Kaggle and Zindi Africa (Nureni & Adekola, 2022; Akinmoluwa et al., 2024), which may not accurately reflect the financial behaviors of Nigerian borrowers using digital lending platforms. Additionally, while previous studies have examined demographic and employment, related predictors, there is limited research incorporating behavioral and transactional data specific to online loan apps, which operate under different lending conditions compared to conventional financial institutions.

Furthermore, while machine learning techniques have been increasingly employed for loan default prediction (Efekodo et al., 2025; Jahanzaib et al., 2024), most studies emphasize algorithmic performance rather than the interpretability of these models in real, world applications. Many existing models focus on accuracy metrics without adequately assessing their practical utility for lenders in Nigeria, where financial literacy, economic instability, and alternative credit scoring mechanisms significantly impact loan repayment behavior. The limited integration of borrower, specific behavioral patterns, such as loan repayment history, frequency of borrowing, and digital financial transactions, presents a crucial gap in effectively predicting defaults in Nigeria's online lending space.

Additionally, while sectoral and macroeconomic factors have been explored in traditional banking contexts (Adewole et al., 2016), research on their influence in digital lending remains sparse. Most studies overlook the role of regulatory frameworks, fintech lending policies, and alternative credit assessment techniques in shaping default risk within Nigeria's online lending market. Addressing these gaps by developing predictive models tailored to online loan apps, incorporating behavioral insights, and balancing model accuracy with interpretability will significantly enhance risk management strategies for digital lenders in Nigeria, hence this study was conducted.

Tomomewo, et al. (2023) investigated the relationship between credit risk management practices and non, performing loans (NPLs) at Nigerian Deposit Money Banks. They employed panel regression (OLS, fixed and random effects) on data from 14 listed banks covering 2013–2022, examining how Capital Adequacy Ratio (CAR), Loan Loss Provisions (LLP), Loan Advances (LA), and Loan, to, Total Assets (LTAR) influence NPL ratios. Results showed a significant positive effect of CAR on NPL/TLR, suggesting that bank buffer levels while meant to protect against risk may signal lax credit control when too high. Their recommendations around tailored provisioning for high, risk categories are directly relevant to improving loan recovery in digital lenders using ML frameworks.

Chukwu et al. (2024) conducted an empirical analysis on how NPLs impact Nigerian bank performance using 2024 data. Applying econometric techniques to assess profitability metrics, they determined that rising NPLs significantly undermine returns. Their study underscores the critical financial implications of non, repayment trends, reinforcing the need for precise default prediction models in digital lending. Their recommendations to strengthen early, warning systems align with your study's objectives in adopting ML for proactive risk control.

Fekadu et al. (2022) evaluated several machine learning models Random Forest, Decision Tree, KNN, SVM, and XGBoost on a non, performing loan dataset from an Ethiopian private bank. They also performed feature importance analysis. XGBoost emerged as the top, performing model on SMOTE, balanced data, and revealed that borrower age, years of employment, and income levels are stronger predictors of default than collateral metrics. This mirrors your interest in demographic and behavioral features in online loan apps.

Pengpeng Yue et al. (2022) examined the effect of digital finance expansion on household debt risk across Sub, Saharan Africa. Their findings revealed that while digital lending fosters financial inclusion, it concurrently increases the likelihood of households falling into debt traps. This insight aids in understanding systemic risks linked to online loan apps in Nigeria, reinforcing the need to adapt digital credit frameworks with predictive, borrower, centred ML models.

Enoch et al. (2021) studied the effects of client appraisal methods on improving lending efficiency in Nigerian microfinance banks (Adamawa State). Using regression on survey data from credit officers, they concluded that improved borrower screening significantly raises repayment performance. Although not ML, based, their scores, based finding supports the concept that enhanced borrower profiling, whether via appraisal forms or behavioural data, boosts predictive accuracy in risk models.

Lu et al. (2024) employed a hybrid k, means clustering + AdaBoost ML model to predict loan default among farmers, investigating the moderating role of financial literacy. Published in *Borsa Istanbul Review*, their results highlighted that borrowers with higher financial literacy are less risk, prone, and clustered behavioural patterns further refine model predictions. These findings support your focus on incorporating app user behaviour and financial knowledge into predictive modelling.

Hassan Bature et al. (2023) proposed a machine learning, based credit, default prediction system using data from Nigerian microloan platforms. Implementing random forest and Naive Bayes, they built a prototype model to detect high, risk borrowers before loan issuance. Their study emphasizes the practical applicability of ML for automating underwriting and credit decisions in real time, mirroring your study's orientation toward digital lenders.

Kokonya & Kimutai (2024) investigated digital lenders in Kenya, focusing on risk identification, quantification, and loan performance among 108 digital lenders. Using correlational analysis, they showed that systematic risk measurement significantly enhances loan performance. They stressed the need for institutional oversight and borrower monitoring, findings that align with your intent to balance predictive accuracy with interpretability and regulatory relevance.

Ogunmokun et al. (2024) assessed debt finance access in Sub, Saharan Africa with a risk, based framing. Using mixed, methods and focusing on SMEs, they examined how data sparse environments impede risk assessment. Although not purely empirical on loan defaults, their insights into sectoral informality and data constraints provide crucial context for understanding modelling challenges in Nigeria's digital lending sphere.

2. MATERIALS AND METHOD

The research adopts a numerical analysis framework, drawing on firsthand data gathered via a well, organized survey instrument to investigate the risk of default in Nigeria's online lending space. To estimate the chances of borrowers failing to repay, the study applies a logistic regression technique, incorporating variables tied to users' economic status, personal characteristics, and behavioral patterns. The target population includes individuals who have borrowed from digital lending platforms, with a stratified random sampling technique ensuring diverse representation. Borrowers from different income levels, employment statuses, and borrowing histories are included, with a minimum sample size of 400 respondents to enhance statistical reliability. The study captures perspectives from both urban and rural areas to reflect varying borrowing behaviors.

Data collection is conducted through a structured questionnaire distributed digitally via Google Form. The questionnaire includes both closed, ended and Likert, scale questions to obtain precise financial and demographic details from respondents. The logit regression model is applied to analyze repayment behavior, categorizing borrowers as either defaulters or non, defaulters based on whether they have ever delayed repayment beyond the due date. This modeling approach helps identify key factors influencing loan repayment and the likelihood of default, providing insights into online lending risks.

To predict the likelihood of loan default (Y), we specify a binary logistic regression model as follows:

$$P(Y=1)=\frac{e^{(y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6+\beta_7X_7+\beta_8X_8)}}{1+e^{y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6+\beta_7X_7+\beta_8X_8}}$$

$$y=\begin{cases}1, & if\ the\ respondent\ has\ ever\ paid\ a\ day\ or\ more\ after\ the\ loan\ due\ date\ (default)\\ 0, & if\ the\ respondent\ has\ always\ paid\ on\ or\ before\ the\ due\ date\end{cases}$$

The model include the following independent variables:

- X_1 = Borrower's Age
- X_2 = Employment Status (1 = Employed, 0 = Unemployed)
- X_3 = Monthly Income (in Naira)
- X_4 = Loan Amount Requested (in Naira)
- X_5 = Loan Duration (in months)
- X_6 = Number of Active Loans
- X_7 = Interest Rate (%)

X_8 = Location of borrower (rural = 0, urban = 1)

3. RESULT AND DISCUSSION

The result of the analysis of data using R programming are presented below in this section.

3.1 Results

Table 1: *Logistic Regression Results*

Predictor	Estimate	Std. Error	z Value	Pr(>|z|)	Odds Ratio (OR)
Intercept	, 4.521	1.158	, 3.903	9.51e, 05	0.0109
Borrower Age	0.07091	0.01904	3.725	0.000196	1.0735
Employment Status	0.6503	0.3400	1.913	0.055802	1.9162
Monthly Income	, 8.125e, 06	3.484e, 06	, 2.332	0.019676	0.99999
Loan Amount	2.785e, 06	8.883e, 07	3.135	0.001716	1.00000
Loan Duration	0.02692	0.01038	2.593	0.009520	1.0273
Active Loans	0.5237	0.1121	4.670	3.01e, 06	1.6883
Interest Rate	0.09095	0.02904	3.132	0.001734	1.0952
Location	0.6866	0.3535	1.943	0.052073	1.9870

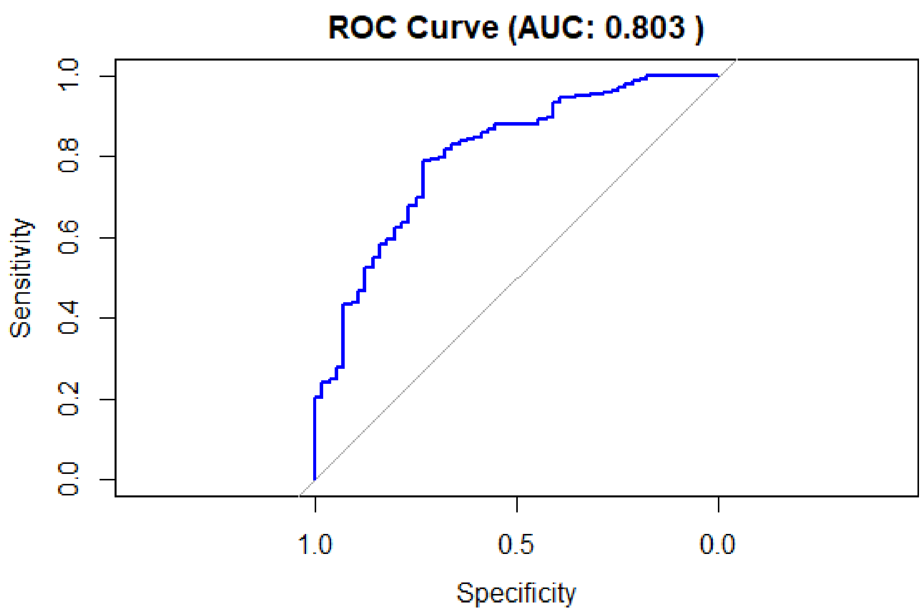

Figure 1: *ROC for AUC*

Table 2: *Model's performance metrics:*

Metric	Value
Accuracy	87.9%
Precision	99.4%
Recall	88.1%
Specificity	83.3%
F1, Score	93.4%

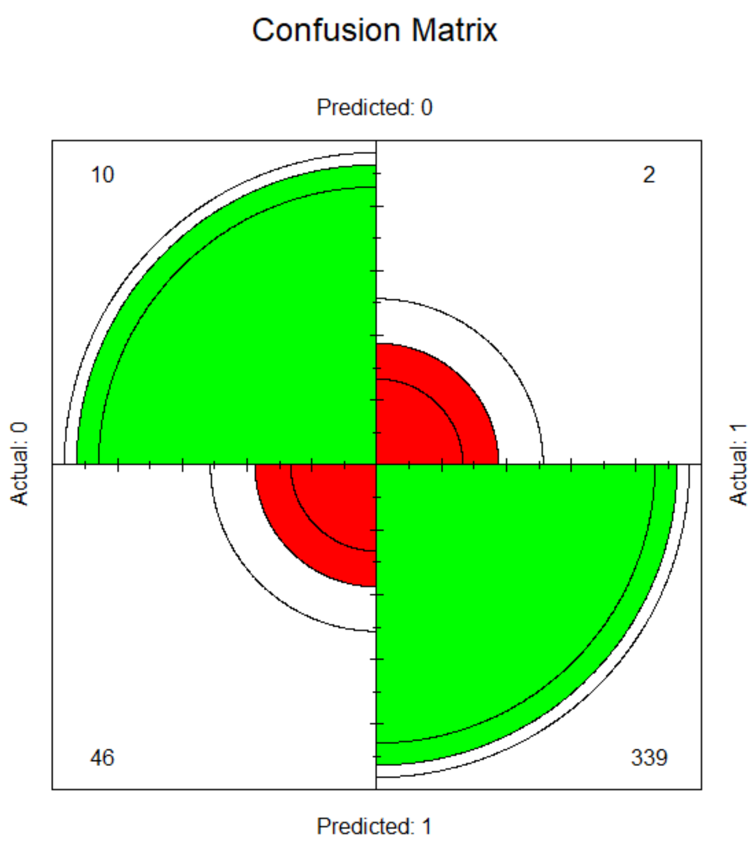

Figure 2: *Confusion Matrix*

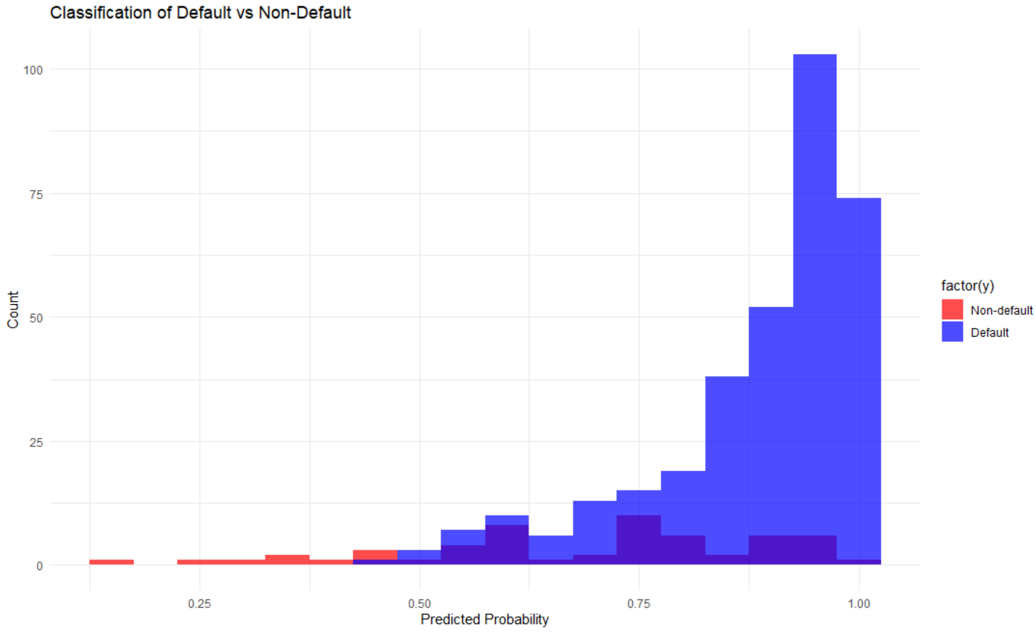

Figure 3: *Loan default classification*

3.2 Discussion

The findings of this study lend strong support to the existing body of literature on loan default prediction, particularly in the context of emerging markets and digital lending systems.

First, the significant positive effect of borrower age on default risk (odds ratio = 1.073; $p < 0.001$) aligns with the results of Fekadu et al. (2022), who found that age was a key predictor of loan default in their Ethiopian study. The implication in both studies is that as borrowers age, financial obligations may increase, such as dependents or debt commitments, leading to greater risk of default. Similarly, Mantey et al. (2021) considered age in their Cox model, although they found it statistically insignificant. The discrepancy may arise from differences in data structure (Cox vs. Logistic model) and borrower segments (community bank vs. digital loans).

The significant negative effect of monthly income on default probability ($p = 0.020$; odds ratio ≈ 0.99999) also corroborates findings from Efekodo et al. (2025) and Nureni & Adekola (2022). These authors established that income is inversely related to loan default and remains one of the most influential features in their machine learning models. This supports the notion that financial capacity, even marginally increased, improves repayment behavior. Similarly, Fekadu et al. reported that income was among the top predictors when building their XGBoost model.

The positive and significant effect of loan amount requested (p = 0.002) is consistent with insights from Adewole et al. (2016), who highlighted that larger loans carry higher default risk, particularly in certain economic sectors. Larger loans may indicate higher financial need or overstretched borrowing limits, both of which can contribute to repayment challenges, especially in high, interest or short, duration environments typical of digital lending platforms.

The study's finding that longer loan durations increase default likelihood (p = 0.010; odds ratio = 1.027) also reinforces findings from Awuza et al. (2022), who showed that repayment timing significantly affects predictive model accuracy. Their study revealed that models incorporating early repayment behavior—an indirect proxy for repayment discipline over time—outperformed others. Longer durations may expose borrowers to financial instability over time, consistent with the rationale in both studies.

Furthermore, the strong positive and significant effect of number of active loans (p < 0.001; odds ratio = 1.688) mirrors the findings of Bature et al. (2023) and Jahanzaib et al. (2024), who emphasized that multiple concurrent credit exposures increase default risk. This is because financial strain intensifies with more repayment commitments, thereby raising delinquency probability—especially when unregulated by centralized databases, as is common in Nigeria's digital lending ecosystem.

The finding that interest rate significantly increases default risk (p = 0.002; odds ratio = 1.095) aligns with the work of Yue et al. (2022), who found that digital finance models offering high, interest, short, term loans can lead to a cycle of over, indebtedness. High interest rates, particularly in the online loan app space, may burden borrowers beyond manageable limits, contributing directly to rising default levels.

The non, significant role of employment status (p = 0.056) contrasts slightly with Mantey et al. (2021), who found employer type (especially public vs. private) to be significant. However, this divergence may be due to differences in how employment is operationalized in traditional versus app, based lending environments. In the digital space, income flow may matter more than formal employment status, especially since many users are self, employed or informal workers.

The marginal insignificance of location (p = 0.052; odds ratio = 1.987) hints at a contextual relevance discussed by Tomomewo et al. (2023) and Kokonya & Kimutai (2024), who addressed regional and demographic disparities in loan performance. While this study suggests urban borrowers are more likely to default, possibly due to higher living costs, the effect isn't strongly significant, warranting further analysis in future studies using more location, sensitive features (e.g., cost, of, living index, urban density).

The ROC AUC value of 0.8029 supports model reliability and aligns with performance metrics reported by Efekodo et al. (2025), Nureni & Adekola (2022), and Akinmoluwa et al. (2024). These studies also reported ML models with accuracy and AUC values ranging from 77% to 83%, confirming that models with moderate, to, high AUC values can serve as dependable tools in real, time loan screening processes.

Figure 2 presents the confusion matrix, which assesses the effectiveness of the model in forecasting loan defaults. It is made up of four essential components: true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP). The model correctly identified 10 borrowers who did not default (TN) and 339 borrowers who defaulted (TP). However, it misclassified 2 borrowers as non, defaulters when they actually defaulted (FP) and incorrectly predicted that 46 borrowers would default when they did not (FN).

The model's overall accuracy in table 2 is approximately 87.9%, indicating that it correctly classified most cases. The precision for identifying defaulters is 99.4%, meaning that when the model predicts a borrower will default, it is correct 99.4% of the time. The recall (true positive rate) is 88.1%, showing that the model captures a high proportion of actual defaulters. Additionally, the specificity (true negative rate) is 83.3%, meaning that 83.3% of non, defaulters were correctly identified. The F1, score, which balances precision and recall, is 93.4%, demonstrating strong predictive capability.

The model accuracy of 0.8791 (or 87.91%) shows that the data represented in figure 3 indicates that the logistic regression model correctly classified 87.91% of the total cases in the dataset. This means that out of all borrowers analyzed, the model was able to accurately predict whether they would default or not in nearly 88 out of 100 cases. A high accuracy score suggests that the model is performing well overall.

4. CONCLUSION

Based on the study findings, the high precision suggests that the model is highly reliable in identifying borrowers at risk of default, which is beneficial for lenders aiming to minimize financial losses. The relatively high recall ensures that most defaulters are captured, but the presence of 46 false negatives suggests that some borrowers who could have been granted loans were mistakenly predicted to default. The significant predictors highlight key financial and demographic factors influencing borrowers' repayment behavior.

To further improve the model, adjustments such as fine, tuning the decision threshold or incorporating additional predictive factors like credit history or spending behavior could help reduce the number of false negatives. While the model is already effective in identifying defaulters, enhancing its recall would ensure even better risk management for lenders, reducing both financial exposure and missed lending opportunities.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Adewole, C., Otubor, C., Mang, N., Ayeni, A., Kairo, I., & Okeke, A. (2016). Investigating the real triggers of loan default in Nigeria. *International Journal of Managerial Studies and Research, 2*(1), 291, 312.

[2] Akinmoluwa, A. A., Odeajo, I., Jimoh, Y., & Afolabi, M. E. (2024). Predicting credit risk: Using machine learning algorithms to predict the creditworthiness of borrowers and determine their likelihood of defaulting on loans in Nigeria. *International Journal of Recent Engineering Science, 10*(2), 46, 53. https://doi.org/10.14445/23497157/IJRES, V10I2P107

[3] Anam, S., Habeebullah, Z., Aftab, S. M. A., Mohamad, M., Asadullah, A., & Syeda, M. (2024). Loan default prediction using machine learning algorithms: A systematic literature review 2020, 2023. *Pakistan Journal of Life and Social Sciences, 22*(2), 6234, 6253. https://doi.org/10.57239/PJLSS, 2024, 22.2.00469

[4] Awuza, A. E., Bello, H., Ahmad, A. A., & Suleiman, M. B. (2022). Default prediction for loan lenders using machine learning algorithms. *Sule Lamido University Journal of Science & Technology, 5*(1), 1, 12. https://doi.org/10.56471/slujst.v5i.222

[5] Bature, H., Musa, M. S., & Adejoh, P. (2023). A machine learning approach to loan default prediction in Nigerian microfinance institutions. *SN Computer Science, 4*(2), 165. https://doi.org/10.1007/s42979, 023, 01832, 2

[6] Chukwu, K. E., Muritala, A. A., Akande, J. A., & Adekunle, T. B. (2024). Impact of non, performing loans on profitability of deposit money banks in Nigeria. *International Journal of Finance & Banking Studies, 13*(1), 88–104. https://doi.org/10.20525/ijfbs.v13i1.1541

[7] Efekodo, K. O., Akinola, O. S., & Waheed, A. A. (2025). Evaluation of machine learning, based algorithm to predicting loan default in Nigeria. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR), 13*(1), 118, 129.

[8] Enoch, O. A., Arabo, D., & Digil, Y. T. (2021). Client appraisal and loan performance in Nigerian microfinance banks: Evidence from Adamawa State. *African Journal of Economic Policy, 28*(2), 59–74. https://journals.unilag.edu.ng/index.php/ajep/article/view/1981

[9] Eweoya, I. O., Adebiyi, A. A., Azeta, A. A., Chidozie, F., Agono, F. O., & Guembe, B. (2019). A Naive Bayes approach to fraud prediction in loan default. *Journal of Physics: Conference Series, 1299*, 012038. https://doi.org/10.1088/1742, 6596/1299/1/012038

[10] Fekadu, M., Geda, T., & Hailu, A. (2022). Loan default prediction using machine learning techniques: Evidence from a private bank in Ethiopia. *Journal of Big Data, 9*(1), 115. https://doi.org/10.1186/s40537, 022, 00629, 7

[11] Jahanzaib, A., Imtiaz, A., & Kehkashan, N. (2024). Advancing financial resilience: A systematic review of default prediction models and future directions in credit risk management. *Heliyon, 10*, e39770. https://doi.org/10.1016/j.heliyon.2024.e39770

[12] Kokonya, J. M., & Kimutai, D. C. (2024). Risk management practices and loan performance among digital lenders in Kenya. *African Journal of Business Management, 18*(2), 36–48. https://doi.org/10.5897/AJBM2023.9634

[13] Lu, Y., Li, X., & Wu, Q. (2024). Financial literacy, risk attitudes, and default risk prediction using machine learning: Evidence from rural China. *Borsa Istanbul Review, 24*(1), 88–102. https://doi.org/10.1016/j.bir.2023.10.005

[14] Mantey, G. K., Dadzie, S. E., Gyasi, I., & Tabi, F. (2021). Application of survival model to analyze default rates of personal bank loans: The case of a bank in Ghana. *African Journal of Management Research (AJMR).* https://dx.doi.org/10.4314/ajmr.v28i2.5

[15] Nureni, A. A., & Adekola, O. E. (2022). Loan approval prediction based on machine learning approach. *FUDMA Journal of Sciences, 6*(3), 41, 50. https://doi.org/10.33003/fjs, 2022, 0603, 830

[16] Ogunmokun, G., Mafimisebi, T. E., & Obembe, O. (2024). Access to debt finance in Sub, Saharan Africa: Risk, based perspectives and implications for SMEs. *Journal of Small Business and Enterprise Development, 31*(1), 91–109. https://doi.org/10.1108/JSBED, 09, 2023, 0354

[17] Tomomewo, S. O., Falayi, R. F., & Uhuaba, O. R. (2023). Credit risk management practices and non, performing loans in Nigerian deposit money banks. *Journal of Risk and Financial Management, 16*(1), 45. https://doi.org/10.3390/jrfm16010045

[18] Yue, P., Wang, R., & Zhu, Y. (2022). Digital finance, household debt and risk in Sub, Saharan Africa. *Journal of African Business, 23*(3), 475–492. https://doi.org/10.1080/15228916.2021.1967648