American Journal of Data Mining and Knowledge Discovery
2026, Vol. 11, No. 2, pp. 18–22
https://doi.org/10.11648/j.ajdmkd.20261102.11

Research Article

A Comparative Evaluation of Machine Learning Algorithms for Anomaly Detection in Mobile Money Transactions

Nnanna Ekedebe[*]

Department of Information Technology, Federal University of Technology, Owerri, Nigeria

Abstract

Nigeria's mobile money ecosystem has expanded rapidly, with mobile money operators processing tens of trillions of naira annually and serving over 220 million subscribers; however, financial institutions lost ₦52.26 billion to fraud in 2024 alone, and automated fraud detection approaches remain largely unexplored specifically within the Nigerian mobile money context. This study presents a comparative evaluation of four machine learning algorithms Logistic Regression, Decision Tree, Random Forest, and XGBoost for anomaly detection in mobile money transactions, with the aim of identifying a viable deployment candidate for Nigerian fintech platforms. The four algorithms were trained and tested on the PaySim synthetic mobile money dataset under identical experimental conditions, using an 80/20 stratified train-test split combined with a hybrid resampling strategy that undersampled the majority class before applying SMOTE to the minority fraud class. Model performance was evaluated using Precision, Recall, F1-Score, AUC-ROC, and Inference Speed. XGBoost achieved the highest F1-Score of 35.70% and AUC-ROC of 99.98%, with all tree-based models recording Recall above 99% and all four algorithms demonstrating real-time viable inference speeds. Precision remained low across the tree-based models, attributed to the synthetic nature of the PaySim dataset and the undersampling strategy applied during training. XGBoost is recommended as the primary deployment candidate for Nigerian fintech platforms, with precision improvement through access to real transaction data and threshold optimization identified as immediate next steps.

Keywords

SMOTE, Nigerian Fintech, PaySim, Random Forest, Anomaly Detection

1. Introduction

With the growth of mobile money in Africa, financial access has been transformed completely, allowing millions to transact without going through the normal banking structures. The need for more automated machine learning solutions for fraud detection in mobile money has become increasingly paramount as fintech platforms across Africa such as OPay, Moniepoint and PalmPay process and move trillions in transactions.

In 2024, Nigerian financial institutions lost a total of ₦52.26 billion to fraud, according to the Nigeria Inter-Bank Settlement System (NIBSS) [1]. This is not surprising, as electronic payment transactions reached a staggering ₦1.07 quadrillion in 2024 which is the first time Nigeria crossed the quadrillion threshold with mobile money operators alone including OPay and PalmPay processing ₦71.5 trillion, serving over 220 million mobile subscribers [1]. Therefore, outdated and manual methods for fraud detection can no longer compete

[*]Correspondence: Nnanna Ekedebe (nnanaekedebe@gmail.com)

Received: 12 July 2026; **Accepted:** 23 July 2026; **Published:** 17 August 2026

with the scale of this development.

Mobile money differs from credit card fraud in transaction types, patterns and even threat actor behaviours. There is also a massive class imbalance, where fraudulent transactions represent only a tiny fraction of total transaction volume. With any delay in fraud detection causing funds to be moved irreversibly, a real-time detection system is required [2].

Machine learning enables automated pattern recognition in large transaction datasets, with algorithms such as Random Forest and XGBoost demonstrating strong performance in fraud detection tasks. These can be deployed for real-time monitoring within transaction pipelines [3].

Most existing studies focus on credit card fraud and not on mobile money transactions, with few comparative studies framing algorithm evaluation around the Nigerian fintech ecosystem. The deployment implications for fintech platforms such as OPay, PalmPay and Moniepoint remain largely unaddressed in the literature.

This study sets out to evaluate four machine learning algorithms for mobile money anomaly detection, trained and tested on PaySim which is a publicly available synthetic mobile money dataset modelled on a real African service. The models are evaluated using metrics such as Precision, Recall, F1-Score, AUC-ROC and Inference Speed, and deployment implications as they relate to the Nigerian fintech fraud prevention landscape are discussed.

2. Related Work

This section reviews literature from traditional rule-based fraud detection through classical machine learning and ensemble methods. These methods made progress, but most do not tackle the Nigerian deployment context or evaluate algorithms specifically on mobile money fraud data gaps which this study aims to address.

Earlier forms of fraud detection relied heavily on predefined rules and thresholds. They were limited by their inability to adapt to ever-evolving fraud patterns, and the high false positive rates and significant operational overheads incurred remained a persistent limitation [4].

As time went on machine learning began to improve detection by learning patterns directly from historical transaction data. Algorithms such as Logistic Regression and Decision Trees were established as early baselines in the literature, with studies demonstrating improved precision over rule-based approaches [5].

Random Forest and XGBoost emerged with strong performances on imbalanced fraud datasets, with SMOTE and other resampling techniques demonstrating their ability to improve recall on minority fraud classes. F1-Score and AUC-ROC were also established as the more appropriate metrics for imbalanced fraud detection tasks [6]. These resampling-based approaches have since been validated widely in credit card fraud research: SMOTE combined with AdaBoost has been shown to strengthen recall on minority fraud classes without excessive precision loss [7], while ensemble and neural approaches incorporating engineered features have further improved detection accuracy [8]. Comparative studies benchmarking classical machine learning against deep learning approaches on imbalanced fraud datasets have similarly found tree-based and ensemble methods to outperform simpler baselines [9, 10], with soft voting ensembles offering additional gains in detection performance on skewed class distributions [11].

With the PaySim dataset being introduced as a benchmark for mobile money fraud research, studies using PaySim demonstrated the viability of machine learning for mobile transaction anomaly detection, with the African mobile money fraud context remaining largely unexplored in existing literature [12]. More recent work has begun to address this gap directly: Azamuke et al. examined fraud detection across rich mobile money transaction datasets in the Sub-Saharan African context [13], while Ogwueleka et al. applied a stacked ensemble technique to the PaySim dataset specifically within the Nigerian mobile money setting, achieving strong recall and reduced false-positive rates [14]. However, comparative benchmarking of multiple standalone algorithms with an explicit Nigerian fintech deployment framing, as undertaken in this study, remains absent from this emerging body of work.

With no study bench marking machine learning algorithms with an explicit Nigerian fintech deployment angle, this study comes into the picture to address that gap.

3. Methodology

This study employs a comparative evaluation design approach to the evaluation of four machine learning algorithms, trained and tested under identical experimental conditions. The entire pipeline was run on Google Colab CPU.

The PaySim synthetic mobile money dataset by Lopez-Rojas et al. was used, which contains 6,362,620 transactions, of which 8,213 are fraudulent, representing 0.13% of the total. The dataset includes 11 features covering transaction types, amount, and sender/recipient balances, all modelled on a real African mobile money service [12].

For preprocessing, non-predictive identifier columns such as nameOrig, nameDest and isFlaggedFraud were removed. Transaction type was one-hot encoded across five categories CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER. The data was split into an 80/20 train/test ratio with class distribution kept intact. Given the scale of the dataset and the computational constraints of the experimental environment, the majority class within the training set was reduced to 50,000 instances, after which SMOTE [15], was applied to bring the minority fraud class up to match, producing a balanced training set of 100,000 instances. This hybrid resampling approach, combining majority class undersampling with SMOTE oversampling, has been shown to improve classification performance on imbalanced datasets [14]. The test set was left entirely unaltered.

The four algorithms selected for benchmarking were chosen for the following reasons: Logistic Regression was selected for its speed, interpretability and suitability as a baseline; Decision Tree for its simplicity and relevance to low-resource deployment; Random Forest for its ensemble approach and strength on imbalanced tabular data; and XGBoost for its gradient boosting capability, which is state-of-the-art for tabular fraud detection [5, 16].

The models were evaluated on Accuracy, Precision, Recall, F1-Score, AUC-ROC and Inference Speed, with F1-Score and AUC-ROC treated as the primary metrics due to the severe class imbalance of 0.13%. Inference Speed was also measured in milliseconds per transaction, which is paramount for real-time deployment assessment.

The results of this study were interpreted against Nigerian fintech deployment requirements, with inference speed assessed for its viability in real-time transaction monitoring. Model interpretability was also discussed in the context of regulatory compliance.

4. Results and Discussion

As shown in Table 1, XGBoost achieved the highest performance across the primary metrics.

Table 1. *Model Performance Comparison.*

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)	Inference (ms)
Logistic Regression	95.93	2.88	93.18	5.59	98.78	0.0002
Decision Tree	99.48	19.83	99.09	33.04	99.28	0.0002
Random Forest	99.47	19.54	99.21	32.66	99.94	0.0049
XGBoost	99.54	21.75	99.63	35.70	99.98	0.0019

4.1. Overall Results

with an F1-Score of 35.70% and AUC-ROC of 99.98%, making it the strongest candidate for deployment in the Nigerian fintech context. With all four models achieving a 99%+ Recall score except Logistic Regression, the tree-based models successfully detected virtually all fraudulent transactions.

4.2. Per-Model Analysis

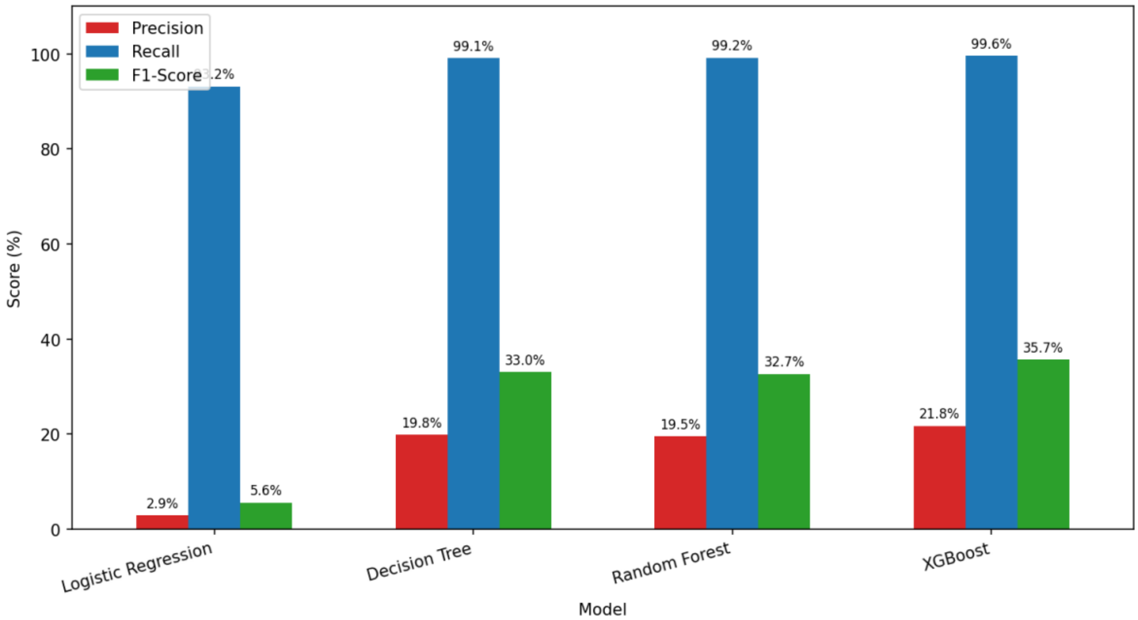

Figure 1. *ML Model Performance Comparison on Precision, Recall and F1-Score across all four models.*

It was observed that Precision collapsed to 2.88% despite an AUC-ROC of 98.78%, which suggests threshold miscalibration. SMOTE creates a 50/50 training distribution, but the real test set is 99.87% legitimate, meaning the default 0.5 cutoff is miscalibrated at evaluation time. This miscalibration is a known linear classifier behavior and not an experimental error. This has direct implications for deployment, as Logistic Regression requires threshold recalibration before it can be considered a viable candidate. As illustrated in Figure 1, the performance gap between Logistic Regression and the tree-based models is evident across all three metrics.

It was observed that All three tree-based models outperformed Logistic Regression significantly on the F1 metric. XGBoost had an edge over Random Forest and Decision Tree on both F1 and AUC-ROC. In the comparison between Random Forest and Decision Tree, both recorded nearly identical F1 scores; however, Random Forest's AUC-ROC of 99.94% compared to Decision Tree's 99.28% suggests better ranking ability. Decision Tree's simplicity makes it relevant for low-resource deployment despite its slightly weaker metrics.

4.3. Inference Speed Analysis

All four models recorded inference speeds between 0.0002ms and 0.0049ms per transaction, suggesting that all are viable for real-time deployment on Nigerian fintech platforms. Logistic Regression and Decision Tree, both at 0.0002ms, are the fastest, which is particularly relevant for ultra-high-volume platforms such as OPay.

4.4. Precision Limitation Analysis

Precision across all three tree-based models remained low at 19–21%, which is attributed to two factors: the synthetic nature of the PaySim dataset, and majority class undersampling reducing the diversity of legitimate transactions seen during training. This was not a methodological flaw but rather a characteristic of the dataset. PaySim remains the field-standard benchmark and the only publicly available mobile money dataset modelled on a real African mobile money service. Access to real transaction data from Nigerian platforms would likely improve precision significantly.

A critical factor contributing to the low precision observed across the tree-based models is the mismatch between the balanced training distribution created through SMOTE and the highly imbalanced distribution of the unaltered test set. Because SMOTE brought the training data to a 50/50 class balance, the default 0.5 classification threshold used at evaluation is not well-suited to the true 0.13% fraud prevalence in the test set, resulting in a large number of false positives despite the models' high Recall and AUC-ROC values. Real-world deployment of these models should therefore not rely on the default probability threshold; instead, threshold optimization, informed by the ROC curve or, given the severity of the class imbalance, the Precision-Recall curve, should be applied to identify a cutoff that better balances Precision and Recall for operational use. This threshold recalibration step, alongside access to real transaction data, is identified as a necessary next step before any of the evaluated models, particularly XGBoost, can be considered production-ready for deployment within the Nigerian fintech ecosystem.

4.5. Deployment Recommendations

Overall, XGBoost is highly recommended as the primary candidate for deployment on fintech platforms given its best F1 and AUC-ROC scores. Decision Tree still remains a viable option for lower-resource environments or edge deployment. Logistic Regression is not recommended unless threshold tuning has been applied. Overall, all models require precision improvement before production deployment, with threshold optimization and feature engineering identified as immediate next steps.

5. Conclusion

XGBoost established itself as the strongest model in this comparative evaluation and is recommended for the Nigerian fintech context. What this study observed as it relates to fintech platforms such as OPay, Moniepoint and PalmPay is that machine learning for real-time mobile money fraud detection is practical today, as demonstrated by the inference speeds recorded across all models. However, precision remains the critical barrier between research and production deployment.

The synthetic nature of the PaySim dataset and the majority class undersampling constraint both impose a ceiling on what these results can be generalised to, and addressing these constraints remains the clearest path forward for this line of research.

Future work should focus more on access to real Nigerian platform transaction data, threshold optimization particularly for Logistic Regression, and feature engineering to close the precision gap. Deep learning approaches should also be explored as the next comparative benchmark, alongside regulatory and explainability considerations relevant to Nigerian fintech compliance.

As Nigeria's digital financial ecosystem and transaction volumes continue to grow, the role of machine learning in securing it has become more critical than ever, making further work in this direction not just valuable but necessary.

Abbreviations

ML	Machine Learning
SMOTE	Synthetic Minority Oversampling Technique
AUC-ROC	Area Under the Receiver Operating Characteristic Curve

NIBSS	Nigeria Inter-Bank Settlement System
PaySim	Payment Simulator (Synthetic Mobile Money Dataset)

Acknowledgments

I acknowledge the support provided by Federal University of Technology Owerri (FUTO), through the Information Communication Technology (ICT) research center.

Author Contributions

Nnanna Ekedebe: Conceptualization, Writing – review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Nigeria Inter-Bank Settlement System (NIBSS), Industry Fraud Report 2024, NIBSS, Feb. 2025. Available: https://nibss-plc.com.ng

[2] M. E. Lokanan, "Predicting mobile money transaction fraud using machine learning algorithms," Applied AI Letters, vol. 4, e85, 2023. https://doi.org/10.1002/ail2.85

[3] P. Alumona, O. Lawal, M. O. Ikhifa, D. O. Agbeso, O. Awele, and D. Olukoya, "Fraud detection in financial transactions using machine learning: Insights from the PaySim mobile money dataset," World Journal of Advanced Research and Reviews, vol. 28, no. 3, pp. 382–392, 2025. https://doi.org/10.30574/wjarr.2025.28.3.4058

[4] Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, "Deep learning in financial fraud detection: Innovations, challenges, and applications," Data Science and Management, 2025. https://doi.org/10.1016/j.dsm.2025.08.002

[5] J. K. Afriyie, K. Tawiah, W. A. Pels, S. Addai-Henne, H. A. Dwamena, E. O. Owiredu, S. A. Ayeh, and J. Eshun, "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," Decision Analytics Journal, vol. 6, p. 100163, 2023. https://doi.org/10.1016/j.dajour.2023.100163

[6] L. Andrade-Arenas and C. Yactayo-Arias, "Comparative analysis of machine learning models for credit card fraud detection using SMOTE for class imbalance," International Journal of Safety and Security Engineering, vol. 15, no. 5, pp. 893–901, 2025. https://doi.org/10.18280/ijsse.150504

[7] E. Ileberi, Y. Sun, and Z. Wang, "Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost," IEEE Access, vol. 9, pp. 165286–165294, 2021. https://doi.org/10.1109/ACCESS.2021.3134330

[8] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A neural network ensemble with feature engineering for improved credit card fraud detection," IEEE Access, vol. 10, pp. 16400–16407, 2022. https://doi.org/10.1109/ACCESS.2022.3148298

[9] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," IEEE Access, vol. 10, pp. 39700–39715, 2022. https://doi.org/10.1109/ACCESS.2022.3166891

[10] G. Airlangga, "Comparative analysis of machine learning models for credit card fraud detection in imbalanced datasets," Journal of Computer Networks, Architecture and High Performance Computing, vol. 6, no. 2, pp. 858–866, 2024. https://doi.org/10.47709/cnahpc.v6i2.3816

[11] M. Azim Mim, N. Majadi, and P. Mazumder, "A soft voting ensemble learning approach for credit card fraud detection," Heliyon, vol. 10, no. 3, e25466, 2024. https://doi.org/10.1016/j.heliyon.2024.e25466

[12] E. A. Lopez-Rojas, A. Elmir, and S. Axelsson, "PaySim: A financial mobile money simulator for fraud detection," in Proc. 28th European Modelling and Simulation Symposium (EMSS), Larnaca, Cyprus, 2016, pp. 249–255. Available: https://www.kaggle.com/datasets/ealaxi/paysim1

[13] D. Azamuke, M. Katarahweire, and E. Bainomugisha, "Financial fraud detection using rich mobile money transaction datasets," in Proc. AFRICOMM 2023, Part II, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 588, Springer, Cham, 2025, pp. 190–208. https://doi.org/10.1007/978-3-031-81573-7_16

[14] F. N. Ogwueleka, A. Imam, and O. Adebisi, "Machine learning approaches for fraud detection in financial systems," International Journal of Advances in Scientific Research and Engineering (IJASRE), vol. 11, no. 9, pp. 1–24, 2025. https://doi.org/10.31695/IJASRE.2025.9.1

[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002. https://doi.org/10.1613/jair.953

[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785