Simion Mbumbu

Elsevier_s_CAS_LaTeX_Single_Column_Template.pdf

Strathmore University (Main Account)

Document Details

Submission ID
trn:oid:::2945:338128523

Submission Date
Dec 18, 2025, 2:00 PM GMT+3

Download Date
Dec 18, 2025, 3:45 PM GMT+3

File Name
Elsevier_s_CAS_LaTeX_Single_Column_Template.pdf

File Size
437.9 KB

14 Pages

7,998 Words

50,165 Characters

30% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

101 Not Cited or Quoted 18%
Matches with neither in-text citation nor quotation marks

69 Missing Quotations 11%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

15% Internet sources
14% Publications
27% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

101 Not Cited or Quoted 18%
Matches with neither in-text citation nor quotation marks

69 Missing Quotations 11%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

15% Internet sources
14% Publications
27% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	University of Sydney on 2024-10-08 — 1%
2	Internet	www.mdpi.com — 1%
3	Internet	arxiv.org — <1%
4	Submitted works	Heriot-Watt University on 2025-08-11 — <1%
5	Submitted works	BB9.1 PROD on 2025-11-30 — <1%
6	Submitted works	University of Surrey on 2025-09-03 — <1%
7	Submitted works	Liverpool John Moores University on 2025-08-05 — <1%
8	Submitted works	University of Hertfordshire on 2024-09-02 — <1%
9	Submitted works	University of Ulster on 2025-12-12 — <1%
10	Internet	assets-eu.researchsquare.com — <1%

11 Internet

dergipark.org.tr <1%

12 Publication

Poonam Nandal, Mamta Dahiya, Meeta Singh, Arvind Dagur, Brijesh Kumar. "Pro... <1%

13 Internet

www.preprints.org <1%

14 Publication

Amjad Iqbal, Rashid Amin. "Time Series Forecasting and Anomaly Detection Usin... <1%

15 Submitted works

Glasgow Caledonian University on 2024-01-14 <1%

16 Submitted works

Liverpool John Moores University on 2025-05-13 <1%

17 Internet

etasr.com <1%

18 Submitted works

jku on 2025-03-26 <1%

19 Submitted works

Mansoura University on 2025-09-10 <1%

20 Submitted works

Strathmore University (Main Account) on 2025-11-28 <1%

21 Submitted works

Uganda Christian University on 2025-05-12 <1%

22 Submitted works

RDI Distance Learning on 2025-05-08 <1%

23 Submitted works

Sim University on 2020-10-22 <1%

24 Submitted works

The Indian Institute Of Management And Engineering Society on 2025-12-08 <1%

25 Internet
pmc.ncbi.nlm.nih.gov <1%

26 Internet
techagro.org <1%

27 Submitted works
University of Hertfordshire on 2025-11-03 <1%

28 Submitted works
University of Houston, Downtown on 2025-07-01 <1%

29 Submitted works
University of Sunderland on 2024-12-12 <1%

30 Internet
ijecs.in <1%

31 Internet
www.ijrah.com <1%

32 Publication
Taiwo O. Soetan, Emmanuel Mogaji. "Financial Services in Nigeria", Springer Scie... <1%

33 Submitted works
Maastricht University on 2025-06-20 <1%

34 Submitted works
University of Central Lancashire on 2025-11-12 <1%

35 Submitted works
University of Westminster on 2023-10-26 <1%

36 Internet
eprints.utar.edu.my <1%

37 Submitted works
Roehampton University on 2025-09-19 <1%

38 Submitted works
University of Essex on 2025-12-12 <1%

39 Submitted works
University of Reading on 2025-06-13 <1%

40 Submitted works
University of Teesside on 2024-05-13 <1%

41 Internet
ijses.com <1%

42 Internet
ijsrem.com <1%

43 Internet
journal-innovations.com <1%

44 Submitted works
Australian Institute of Higher Education on 2025-08-05 <1%

45 Submitted works
Liverpool John Moores University on 2025-05-15 <1%

46 Submitted works
Strathmore University (Main Account) on 2025-11-27 <1%

47 Submitted works
University of Essex on 2025-12-12 <1%

48 Submitted works
University of Hertfordshire on 2025-11-03 <1%

49 Submitted works
University of Ulster on 2025-12-08 <1%

50 Internet
kjan.or.kr <1%

51 Internet
publikasi.dinus.ac.id <1%

52 Internet
www.ijirset.com <1%

53	Publication	Girish Kadamathikuttiyil Karthikeyan, Biswajit Bhowmik. "Intelligent money laun... <1%
54	Submitted works	Islington College,Nepal on 2025-12-17 <1%
55	Submitted works	Manchester Metropolitan University on 2025-09-15 <1%
56	Submitted works	University of Hertfordshire on 2024-09-02 <1%
57	Submitted works	University of North Texas on 2025-04-28 <1%
58	Submitted works	University of Wollongong on 2025-11-18 <1%
59	Internet	d197for5662m48.cloudfront.net <1%
60	Internet	ouci.dntb.gov.ua <1%
61	Internet	rjwave.org <1%
62	Internet	www.databridgemarketresearch.com <1%
63	Internet	www.ewadirect.com <1%
64	Submitted works	UCL on 2025-10-27 <1%
65	Submitted works	University of Huddersfield on 2024-01-22 <1%
66	Submitted works	University of Northumbria at Newcastle on 2023-07-11 <1%

67 Submitted works
University of Wolverhampton on 2025-08-22 <1%

68 Internet
corz.org <1%

69 Internet
hal.science <1%

70 Internet
isjem.com <1%

71 Internet
theses.fr <1%

72 Internet
www.igi-global.com <1%

73 Internet
www.multiresearchjournal.com <1%

74 Submitted works
Coventry University on 2025-04-10 <1%

75 Submitted works
Coventry University on 2025-12-04 <1%

76 Publication
Jing Xian Ooi. "Effective Credit Card Fraud Detection Using Data Mining Techniqu... <1%

77 Publication
Mukungi Mutemi, Paul. "Organized Retail Crimes and Fraud Detection for E-Com... <1%

78 Submitted works
Nottingham Trent University on 2025-09-12 <1%

79 Submitted works
The Robert Gordon University on 2024-08-03 <1%

80 Submitted works
University of Edinburgh on 2025-12-06 <1%

81 Submitted works
University of Northumbria at Newcastle on 2025-05-15 <1%

82 Submitted works
University of Ulster on 2025-10-02 <1%

83 Submitted works
University of Zambia on 2024-05-09 <1%

84 Internet
cdn.clinicaltrials.gov <1%

85 Internet
fastercapital.com <1%

86 Internet
jutif.if.unsoed.ac.id <1%

87 Publication
Ahmad, Maitha. "Predictive Policing - Leveraging CCTV Data and AI for Crime Hot... <1%

88 Publication
Al-Balushi, Abrar Ahmed. "Applying Supervised Machine Learning Algorithms & E... <1%

89 Submitted works
Asia Pacific University College of Technology and Innovation (UCTI) on 2025-06-13 <1%

90 Submitted works
Bocconi University on 2025-11-03 <1%

91 Submitted works
City University College of Science and Technology on 2024-11-24 <1%

92 Submitted works
EC-Council University on 2025-09-11 <1%

93 Submitted works
HELP UNIVERSITY on 2025-11-11 <1%

94 Submitted works
Higher Education Commission Pakistan on 2024-10-29 <1%

95 Submitted works
King's College on 2025-04-24 <1%

96 Submitted works
Mount Kenya University on 2025-10-08 <1%

97 Publication
Nouhaila Hanbali, Ahmed El-Yahyaoui. "Advanced machine learning and deep lea... <1%

98 Publication
Samuel Odoom, Eric Opoku Osei, Enock Quansah Effah, Victoria Boafo, Seyram D... <1%

99 Submitted works
Strathmore University (Main Account) on 2025-12-01 <1%

100 Submitted works
The Robert Gordon University on 2025-09-16 <1%

101 Submitted works
The University of the West of Scotland on 2025-12-07 <1%

102 Submitted works
University of Bradford on 2023-09-06 <1%

103 Submitted works
University of Canberra on 2024-08-22 <1%

104 Submitted works
University of Hertfordshire on 2025-03-03 <1%

105 Submitted works
University of Johannsburg on 2025-05-30 <1%

106 Submitted works
University of Northumbria at Newcastle on 2025-05-14 <1%

107 Submitted works
University of Ulster on 2025-04-17 <1%

108 Submitted works
University of Wales Swansea on 2024-09-30 <1%

109 Submitted works
Zambia Centre for Accountancy Studies on 2025-05-10 <1%

110 Internet
dalspace.library.dal.ca <1%

111 Internet
download.bibis.ir <1%

112 Internet
dspace.unive.it <1%

113 Internet
jisem-journal.com <1%

114 Internet
www.ijraset.com <1%

115 Internet
www.jetir.org <1%

116 Internet
www.jmsrr.com <1%

117 Publication
"Computational Intelligence Techniques for 5G Enabled IoT Networks", Springer ... <1%

118 Submitted works
Dublin Business School on 2025-08-28 <1%

119 Submitted works
Hong Kong University of Science and Technology on 2024-05-19 <1%

120 Submitted works
Makerere University on 2025-12-11 <1%

121 Publication
Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Re... <1%

122 Submitted works
SP Jain School of Global Management on 2024-01-16 <1%

123 Submitted works

Tilburg University on 2024-12-02 <1%

124 Submitted works

University of Bolton on 2025-10-14 <1%

125 Submitted works

University of Hertfordshire on 2023-09-18 <1%

126 Submitted works

University of Stirling on 2025-04-11 <1%

127 Submitted works

University of Witwatersrand on 2025-11-01 <1%

128 Submitted works

Intercollege on 2022-05-08 <1%

129 Submitted works

National Research University Higher School of Economics on 2020-05-25 <1%

130 Submitted works

University of Essex on 2024-09-18 <1%

131 Submitted works

University of Westminster on 2025-11-13 <1%

132 Submitted works

Zambia Centre for Accountancy Studies on 2025-06-27 <1%

Mobile Money Fraud Detection Using Machine Learning: A CRISP-DM Approach⋆,⋆⋆

Mr David Ongaro M.E[a,1] (Researcher)

[c] *Strathmore University, Ole Sangale, Nairobi, Nairobi, Kenya*

ARTICLE INFO

Keywords:
Fraud detection
machine learning
LightGBM
XGBoost
deep learning
imbalanced data
financial transactions
CRISP-DM

ABSTRACT

Financial fraud has emerged as a critical challenge in the digital economy, causing significant losses to financial institutions and consumers worldwide. While machine learning offers promising solutions for fraud detection, limited research addresses the specific constraints of mobile money ecosystems in developing regions, particularly under severe class imbalance and resource limitations. This study presents a comprehensive comparative analysis of five machine learning algorithms—LightGBM, XGBoost, Random Forest, Logistic Regression, and Linear Support Vector Classifier (LinearSVC)—for detecting fraudulent mobile money transactions in East Africa. Following the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, we evaluated these algorithms on a real-world dataset comprising 95,662 transactions with an extreme fraud rate of 0.2% (1:495 imbalance ratio). Our methodology encompassed data understanding, preparation, exploratory data analysis, feature engineering, model training with class weighting, and performance evaluation using imbalance-appropriate metrics. Results demonstrated that LightGBM achieved superior overall performance with the highest precision-recall balance and PR-AUC, followed closely by Random Forest. Notably, XGBoost underperformed relative to its typical benchmark dominance in other fraud detection contexts, highlighting the importance of algorithm evaluation on domain-specific data rather than relying solely on generalized performance trends. Logistic Regression and LinearSVC provided acceptable baseline performance. These findings provide actionable insights for mobile money operators in emerging markets seeking to balance fraud prevention with operational efficiency, demonstrating that algorithm selection must be tailored to specific data distributions, infrastructure constraints, and deployment environments. The study underscores the critical importance of using precision-recall metrics over accuracy when evaluating models on highly imbalanced datasets characteristic of real-world fraud detection scenarios.

1. Introduction

The digital financial revolution has transformed the economic landscape of developing regions, with mobile money services emerging as a critical driver of financial inclusion Chauhan, Uppal, Gupta, Mapari and Saini (2025). Globally, mobile money platforms have enabled over 1.2 billion registered accounts as of 2024, with the highest concentration in Sub-Saharan Africa (GSMA, 2024). East Africa, particularly Kenya, has pioneered this transformation through services like M-Pesa, which has become synonymous with mobile financial services and has lifted millions out of poverty by providing access to formal financial systems.

However, the rapid expansion of mobile money services has created new opportunities for fraudulent activities, resulting in substantial financial losses globally Sariat, Siddique, Hossain, Islam and Rahman (2025). For instance, in the United Kingdom total losses to payment fraud reached £1.17 billion in 2023, with authorised push payment (APP) scams now constituting the largest share of losses UK Finance (2024). On a global scale, payments fraud is estimated at over $200 billion annually McKinsey & Company (2023), underscoring the critical need for effective detection mechanisms. Estimates from the International Monetary Fund (IMF) indicate that cybercrime alone cost the world economy approximately $8.5 trillion in 2023, a figure projected to grow () (IMF). In Kenya alone, 25.9% of mobile money users reported experiencing financial losses due to online fraud in 2023 (Central Bank of Kenya, 2023). These statistics underscore the urgent need for robust fraud detection mechanisms to maintain user trust and ensure the sustainability of mobile money ecosystems Lu (2024).

Fraud detection in mobile money transactions presents unique challenges that distinguish it from traditional financial fraud detection Souran and Shah (2025). First, the severe class imbalance where fraudulent transactions

david.mauti@strathmore.edu (D.O. M.E)
ORCID(s):

constitute the minority class makes it difficult for standard machine learning models to learn fraud patterns effectively Lu (2024); Psychoula, Gutmann, Mainali, Lee, Dunphy and Petitcolas (2021); **?**. Second, fraudsters continuously adapt their tactics, employing sophisticated techniques such as social engineering, SIM swapping, and deepfake impersonation Sariat et al. (2025); Muqattash and Kharbat (2023); Papasavva, Lundrigan, Lowther et al. (2025). Third, the real-time nature of mobile money transactions requires detection systems that can process high volumes of transactions with minimal latency while maintaining high accuracy Liu, Tang, Yang, Zhou and Cha (2025); Arzu, Bajwa, Obaidullah, Waheed, Alam, Ali and Khan (2025).Third, the real-time nature of mobile money transactions requires detection systems that can process high volumes of transactions with minimal latency while maintaining high accuracy Liu et al. (2025); Arzu et al. (2025). Legacy batch-processing methods are obsolete in this context, as modern payment rails demand sub-second fraud risk scoring McKinsey & Company (2023). This necessitates the use of stream processing frameworks and machine learning models capable of millisecond-latency inference on continuous data streams Amazon Web Services (2024); Liu et al. (2025)

Traditional rule-based fraud detection systems struggle to keep pace with evolving fraud tactics Arzu et al. (2025) and generate high false positive rates, leading to legitimate transaction rejections and poor customer experience Lokanan (2023); Lu (2024). There is a critical need for adaptive, intelligent fraud detection systems that can learn complex patterns from historical data, generalize to new fraud types, and operate effectively under severe class imbalance conditions Theodorakopoulos, Theodoropoulou, Tsimakis and Halkiopoulos (2025). Next-generation approaches, such as graph neural networks, are being developed by modelling complex relational structures and detecting anomalies, even in highly skewed data Cheng, Zou, Xiang et al. (2025).

This research aims to develop and evaluate a machine learning-based fraud detection framework for mobile money transactions using the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology. We endeavour to first analyse the characteristics of mobile money transaction data across different product categories and channels. Secondly, investigate and implement effective techniques for handling severe class imbalance. Thirdly, implement and compare multiple machine learning algorithms (Logistic Regression, LinearSVC, Random Forest, XGBoost, and LightGBM) for fraud detection, evaluating their performance using appropriate metrics for imbalanced datasets. Lastly, discuss practical considerations for deploying the fraud detection system in a real-world mobile money environment.

2. Literature Review

2.1. Mobile Money and Fintech Security

Mobile money has revolutionized financial inclusion and profoundly impacted developing economies, where mobile money services (MMS) have emerged as a critical driver of financial inclusion Sariat et al. (2025); Lokanan (2023); Hanbali and El-Yahyaoui (2025). M-Pesa in Kenya serves as the flagship success story connecting millions of previously unbanked individuals GSMA (2024). While these platforms offer convenience, speed, and accessibility, their widespread adoption has simultaneously increased the risk of financial fraud, leading to substantial economic losses globally Souran and Shah (2025); Muqattash and Kharbat (2023). Financial losses due to fraud are projected to increase significantly across various payment channels in the coming years Wang and Yang (2022); Muqattash and Kharbat (2023). The vulnerability of digital finance is significant in high-adoption markets. For instance, in Kenya a staggering 25.9% of mobile money users reported financial losses due to cybercrime in 2023 National Kenya Computer Incident Response Team Coordination Centre (National KE-CIRT/CC) (2023). These numbers underscore the urgent need for robust fraud detection mechanisms to sustain the mobile money ecosystem.

Recent research on mobile money fraud in East Africa has gained momentum, with several studies providing empirical evidence from the region. Azamuke, Katarahweire and Bainomugisha (2025) utilized rich mobile money transaction datasets to detect financial fraud, demonstrating the effectiveness of machine learning in East African contexts. Lokanan (2023) showed that machine learning algorithms can effectively predict mobile money transaction fraud, while Botchey, Qin and Hughes-Lartey (2020) conducted a comprehensive cross-case analysis comparing Support Vector Machines, gradient boosted decision trees, and Naïve Bayes algorithms specifically for mobile money fraud prediction in Sub-Saharan Africa. These foundational studies establish the viability of ML-based approaches in the region. **However, they often rely on static datasets with short observation windows, leaving open questions about concept drift and model robustness against the rapidly evolving fraud tactics typical of informal economies.**

The fraud landscape in mobile money systems is characterized by sophisticated and evolving tactics. Mambina, Ndibwile and Michael (2022) developed machine learning approaches to classify Swahili smishing attacks targeting

mobile money users, highlighting the linguistic and cultural dimensions of fraud in East Africa. Social engineering remains a primary attack vector, with Tagbo, Adekoya and Mensah (2024) proposing machine learning solutions to mitigate mobile money social engineering attacks. The security challenges extend beyond individual fraud to systemic vulnerabilities, including SIM swap attacks, phishing schemes, and increasingly sophisticated AI-driven methods. Neza, Joseph and Joseph (2022) identified the "e-money security dilemma" facing Sub-Saharan Africa, where advanced cybersecurity mechanisms must coexist with legacy mobile payment systems.

The regulatory and infrastructure dimensions are equally critical. Sanni, Akinyemi, Olalere, Olajubu and Aderounmu (2023) developed a predictive cyber threat model specifically for mobile money services, emphasizing the importance of proactive threat intelligence. Mollik and Majeed (2025) examined AI-driven cybersecurity in mobile financial services, focusing on enhancing fraud detection and privacy in emerging markets. These studies underscore that effective fraud detection requires not only technical solutions but also supportive regulatory frameworks and industry collaboration.

However, it is critical to note that the majority of extant fraud detection literature focuses on credit card transactions in developed economies. **These findings often do not transfer directly to the mobile money context in East Africa**, where user behaviors, transaction velocities, regulatory environments, and fraud typologies (such as SIM swaps, social engineering, and agent collusion) differ fundamentally from credit card theft patterns Mambina et al. (2022); Neza et al. (2022). This contextual gap necessitates domain-specific research using real mobile money transaction data from the region.

2.2. Machine Learning Approaches to Fraud Detection

2.2.1. Traditional Machine Learning Algorithms

Machine learning and Artificial Intelligence (AI) have become essential tools for overcoming the shortcomings of conventional rule-based methods, enabling the identification of subtle anomalies that traditional systems overlook Sariat et al. (2025). Random Forest, as an ensemble learning method, has shown particular strength in financial fraud detection due to its resistance to overfitting and ability to handle class imbalance effectively Afriyie, Tawiah, Prah, Annan and Weber (2023); Mambina et al. (2022).Kumar (2024); Sundaravadivel, Isaac, Elangovan, KrishnaRaj, Rahul and Raja (2025) demonstrated enhanced fraud detection in financial transactions using hyperparameter-tuned Random Forests, achieving superior accuracy through systematic optimization. Raturi (2024) conducted a comparative analysis confirming Random Forest's effectiveness in handling high-dimensional data and capturing non-linear fraud patterns.When applied to synthetic mobile money transaction data, the Random Forest Classifier achieved an exceptional Area Under the Precision-Recall Curve (AUPRC) of 0.9998 Sariat et al. (2025).

Support Vector Machines (SVM) have been widely applied, with variable performance depending on kernel selection and data characteristics. Botchey et al. (2020) evaluated SVM effectiveness specifically for mobile money fraud prediction in Sub-Saharan Africa, finding that while useful, SVMs generally show slightly inferior performance compared to ensemble methods in highly imbalanced fraud detection scenarios. **Moreover, the success of these supervised methods relies heavily on the assumption of abundant, high-quality labeled data, a distinct luxury in many operational mobile money environments where varying verification standards lead to significant label noise.**

2.2.2. Gradient Boosting Methods: XGBoost and LightGBM

Gradient boosting techniques, particularly XGBoost and LightGBM, have emerged as dominant approaches in fraud detection literature. The XGBoost model is repeatedly highlighted for its high performance, efficiency, and strong capability in handling imbalanced data Theodorakopoulos et al. (2025); Sariat et al. (2025); Kandi (2025); Azamuke et al. (2025).Hájek, Abedin and Sivarajah (2022) developed an XGBoost-based framework specifically for fraud detection in mobile payment systems, demonstrating the algorithm's superior discrimination capability in identifying fraudulent mobile transactions. **?**Azamuke et al. (2025) found XGBoost to be the most effective algorithm, achieving high Matthews Correlation Coefficient (MCC) scores (up to 0.82) and efficiency metrics suitable for real-time deployment. Al-Asadi, Alissa, Bhushan and Al-Azzawi (2025) showed that XGBoost's effectiveness is further enhanced when combined with advanced data balancing techniques, achieving exceptional results in financial fraud detection. **It is crucial to note, however, that many of these exceptional performance metrics are reported on synthetic datasets (e.g., PaySim). Such results may not fully translate to real-world production environments where data noise and complex feature interactions can degrade theoretical performance.**

Recent innovations have focused on hybrid and ensemble approaches. Kandi (2025) enhanced XGBoost performance by integrating LSTM networks, demonstrating improved fraud detection through the combination of gradient boosting with sequential pattern recognition. Theodorakopoulos et al. (2025) presented a big data-driven approach using XGBoost and CatBoost for scalable credit card fraud detection, highlighting the algorithms' capacity to handle massive transaction volumes in real-time environments.

LightGBM has gained prominence for its superior speed and efficiency, particularly with imbalanced datasets. Zhao, Liu and Zhao (2024a) developed an improved LightGBM approach specifically addressing extremely imbalanced data in credit card fraud detection, demonstrating significant performance gains. Ramesh and K (2025) proposed a stacked LightGBM-XGBoost model with SHAP-based fraud detection, combining the strengths of both algorithms. Zheng, Chen, Cao, Xu, Xing and Jin (2024a); Zheng, Yu, Cao, Xu, Xing and Jin (2024b) developed advanced payment security systems integrating XGBoost, CatBoost, LightGBM, and SMOTE, demonstrating state-of-the-art performance across multiple financial fraud detection tasks.

Lu (2024) specifically investigated improving fraud detection in mobile payments with machine learning ensembles, showing that combining multiple gradient boosting models yields superior results. Renukadevi, Manujakshi, Shashidhar and Sivakumar (2025) demonstrated fraud detection in financial transactions using gradient boost with hybrid optimization, further advancing the gradient boosting paradigm.

2.2.3. Deep Learning and Neural Network Approaches

Deep learning (DL) models are increasingly utilized because they can automatically extract features and detect complex, non-linear relationships in financial data Hanbali and El-Yahyaoui (2025); Chen, Zhao, Xu, Nie and Zhang (2025). Architectures such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are particularly adept at modeling sequential transaction histories Kandi (2025); Sariat et al. (2025). Al-Khasawneh, Faheem, Alsekait, Abubakar and Issa (2025) developed hybrid neural network methods specifically for credit card fraud detection, demonstrating the effectiveness of combining multiple neural architectures. Chen et al. (2025) provided a comprehensive review of deep learning innovations in financial fraud detection, covering challenges and applications across various domains.

Advanced architectures have emerged for specific fraud detection challenges. Kodete (2023) applied mathematical modeling and deep learning to fraud detection in mobile financial transactions, establishing theoretical foundations for neural network applications in mobile money contexts. Silva, Macêdo, Zanchettin, Oliveira and Filho (2021) pioneered multi-class mobile money service fraud detection by integrating supervised learning with adversarial autoencoders, demonstrating the power of generative models in learning complex fraud patterns. Yussif, Takyi, Gyening and Boadu-Acheampong (2025) developed an advanced mobile money fraud detection system using CNN-BiLSTM with optimized with Stochastic Gradient Descent (SGD), achieving a precision of 0.9927, an accuracy of 0.9928, and a recall of 0.9929. through sequential pattern recognition. The literature also documents the increasing utilization of complex hybrid DL models, such as those combining Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), and Multilayer Perceptron (MLP) within a stacking ensemble framework for CCFD Bonde and Bichanga (2025)

Attention mechanisms have shown particular promise. Dhasaratham, Balassem, Bobba, Ayyadurai and Sundaram (2024) developed an attention-based Isolation Forest integrated ensemble machine learning algorithm for financial fraud detection, demonstrating how attention mechanisms can focus on salient fraud indicators. Wold (2023) examined fraud detection in mobile banking based on artificial intelligence, providing insights into AI deployment in production mobile banking environments.

Despite promising theoretical results from Deep Learning models, **their practical deployment in mobile money contexts raises important operational considerations**. Complex neural architectures may introduce inference latencies exceeding acceptable thresholds for USSD-based real-time transaction processing, particularly in infrastructure-constrained environments common in East African markets Muqattash and Kharbat (2023); Jeyachandran, Akisetty, Subramani, Goel, Singh and Shrivastav (2024). The computational overhead of maintaining and updating deep learning models in resource-limited settings represents a trade-off often under-discussed in academic studies focusing purely on predictive performance metrics.

2.2.4. Hybrid and Ensemble Approaches

Recent research emphasizes that hybrid approaches often outperform single algorithms.Shinde, Chadha and Shitole (2021) demonstrated that detecting fraudulent transactions using hybrid fusion techniques yields superior results compared to individual classifiers. Zhao, Zhang and Zhang (2024b) developed a hybrid ML model that combines

LightGBM and neural networks to tackle skewed data with focal loss, achieving robustness. Jeyachandran et al. (2024) leveraged machine learning for real-time fraud detection in digital payments, demonstrating the practical deployment viability of ensemble approaches. **While hybrid ensembles often report marginal metric gains, the added computational complexity and loss of interpretability may be difficult to justify for operators with limited MLOps capabilities, where maintenance and explainability are paramount.**

2.3. Addressing Class Imbalance in Fraud Detection

Severe class imbalance in fraud datasets is a pervasive methodological constraint, often leading to biased models favoring the majority class. This affects the reliability and robustness of fraud detection, necessitating advanced resampling or weighting techniques AL-Dahasi, Alsheikh, Khan and Jeon (2024); Doddamani, Girish and Bhowmik (2024); Botchey, Qin, Hughes-Lartey and Ampomah (2022).

2.3.1. Oversampling Techniques: SMOTE and Variants

SMOTE (Synthetic Minority Over-sampling Technique) has been extensively researched between 2020-2025, consistently showing significant enhancement of models' ability to detect fraudulent instances Gupta (2025). SMOTE generates synthetic minority class samples based on feature space similarities of nearest neighbors, balancing datasets without simple duplication Ahmed, Axelsson, Li and Sagheer (2025). Recent applications demonstrate SMOTE's effectiveness across diverse fraud detection contexts, leading to improved recall and F1-scores.

Advanced SMOTE variants have emerged to address specific challenges. Bonde and Bichanga (2025) demonstrated that SMOTE-ENN (SMOTE-Edited Nearest Neighbor), which combines oversampling with noise reduction. SMOTE-ENN is consistent and stable with better precision-recall balance when combined with ensemble deep learning models Ahmed et al. (2025). Zheng et al. (2024b,a) integrated SMOTE with XGBoost, CatBoost, and LightGBM in advanced payment security systems, achieving exceptional results in highly imbalanced fraud detection scenarios. **However, many studies utilizing SMOTE fail to account for the risk of identifying synthetic artifacts rather than genuine fraud patterns, particularly when the minority class is extremely sparse. In contrast, algorithm-level approaches like class weighting preserve the integrity of the original feature space.**

Albalawi and Dardouri (2025) showed how combining traditional and deep learning models with class imbalance mitigation techniques significantly improves fraud detection performance. Their comprehensive study established best practices for SMOTE application across different algorithm families. Gupta (2025) developed an enhanced framework using robust feature selection with stacking ensemble models specifically designed for imbalanced fraud datasets, demonstrating the synergy between feature engineering and resampling techniques.

2.3.2. Algorithm-Level and Threshold Optimization Approaches

Beyond data-level techniques, algorithm-level approaches have proven effective. Techniques include Cost-Sensitive Learning (CSL), which assigns higher costs or weights to the misclassification of minority class instances, forcing the model to prioritize fraud detection Chen, Duan, Kang and Qiu (2022); Zhao et al. (2024a). Doddamani et al. (2024) developed money laundering detection in imbalanced e-wallet transactions with threshold optimization, demonstrating that adaptive thresholds significantly improve detection rates while managing false positives. Furthermore, Focal Loss is a specialized loss function often used in deep learning, designed to down-weight easily classified samples and focus training effort primarily on difficult-to-classify fraudulent instances Albalawi and Dardouri (2025). The application of Decision Threshold Adjustment (DTA) allows fine-tuning the model's final output probability threshold to align with specific operational demands, balancing the critical trade-off between false positives and false negatives Chen et al. (2022). Green (2025) examined AI-driven financial intelligence systems incorporating risk detection and strategic analysis, emphasizing the importance of cost-sensitive learning in production fraud detection systems. ?Albalawi and Dardouri (2025) established that accuracy is misleading when classes are imbalanced, and that precision-recall plots are more informative than ROC plots for highly imbalanced fraud detection datasets.

A critical limitation in existing literature is the inappropriate use of evaluation metrics. Although earlier works frequently utilized Accuracy and ROC-AUC as primary performance indicators, **these measures can be deceptive in fraud detection contexts** Zhao et al. (2024a); Albalawi and Dardouri (2025). A model can achieve 99.8% accuracy by simply classifying all transactions as legitimate in datasets with 0.2% fraud rates, yet completely fail at its primary objective. Much of the existing literature fails to prioritize Precision-Recall (PR) metrics, which are mathematically more rigorous for measuring minority class detection performance AL-Dahasi et al. (2024). This methodological gap undermines the comparability and practical applicability of many published results.

2.4. Real-Time Fraud Detection and Deployment Considerations

The operational deployment of fraud detection systems introduces challenges beyond model accuracy. Muqattash and Kharbat (2023) addressed mobile payment fraud detection by leveraging machine learning for rapid analysis, demonstrating low-latency prediction capabilities essential for real-time systems.Gupta (2025) examined real-time online payment fraud detection using machine learning algorithms in financial systems, highlighting infrastructure requirements for production deployment. However, both studies focus on relatively well-resourced environments and provide limited guidance on how such architectures can be adapted to infrastructure-constrained mobile money platforms in East Africa, where network instability and legacy systems remain prevalent.

Research strongly advocates for the use of distributed computing platforms, such as PySpark, which are instrumental in enhancing the scalability, computational efficiency, and real-time processing capability of fraud detection systems Theodorakopoulos et al. (2025). These platforms enable parallelized processing and in-memory computation, effectively reducing latency in prediction phases Theodorakopoulos et al. (2025). Kaur (2025) proposed enhancing fraud detection in portable wallet payment systems using machine learning through a hybrid approach, addressing mobile-specific constraints like limited computational power and network latency. Jeyachandran et al. (2024) demonstrated practical implementations of real-time fraud detection in digital payments. Yet, these works rarely quantify the operational trade-offs between sophisticated distributed architectures, the cost and skills to maintain them in production, which may limit their direct applicability for many mobile money operators.

Pattern analysis and behavioral approaches have emerged as complementary strategies. Cochrane, Gomez, Warmerdam, Flores, McCullough, Weinberger and Pirouz (2021) developed pattern analysis methods for transaction fraud detection, showing that temporal and sequential patterns enhance static feature-based detection. Sa'adah and Pratiwi (2020) classified customer actions on digital money transactions using Probabilistic Neural Networks (PNN), demonstrating the value of user behavior modeling. However, most of these behavior-based approaches are evaluated on relatively short time horizons and controlled datasets, leaving open questions about their robustness and adversarial adaptation in real-world mobile money ecosystems.

2.5. Research Gap and Study Positioning

While extensive research exists on fraud detection in credit card transactions and banking systems Afriyie et al. (2023); Dinesh (2024); Raturi (2024), mobile money fraud detection—particularly in East African contexts—remains relatively underexplored. Most studies focus on Western markets with different transaction patterns, regulatory environments, and fraud tactics. Although recent work by Azamuke et al. (2025); Azamuke, Katarahweire, Businge, Kizza, Opio and Bainomugisha (2023); Lokanan (2023); Botchey et al. (2020); **?**); Yussif et al. (2025) has begun to address this gap, comprehensive studies applying state-of-the-art machine learning techniques to real-world mobile money transaction data from East Africa remain limited.

A recurring methodological limitation across the fraud detection literature is the reliance on synthetic datasets (e.g., PaySim, credit card simulation datasets), which often fail to capture the noise, missing data patterns, and complex feature interdependencies found in real-world mobile money transaction logs Sariat et al. (2025); Botchey et al. (2020). Consequently, models optimized on such data may present challenges in generalization when deployed in production environments with real-world noise and evolving fraud patterns. Furthermore, many published studies lack transparency regarding data provenance, limiting reproducibility and practical applicability Emran and Rubel (2024). **As a result, current literature offers limited guidance for operators seeking to build deployable, low-latency fraud systems on real East African transaction logs. Existing frameworks often conflate offline predictive accuracy with production readiness, ignoring the infrastructure constraints of the target deployment environment.**

This research addresses these gaps by implementing and comparing multiple machine learning algorithms (Logistic Regression, LinearSVC, Random Forest, XGBoost, and LightGBM) on actual mobile money platform data from East Africa, employing comprehensive class imbalance mitigation strategies, and providing practical deployment insights specific to this rapidly growing and critically important market. By following the CRISP-DM methodology, this study provides a replicable framework for mobile money operators to develop, evaluate, and deploy fraud detection systems in resource-constrained environments.

3. Methodology

This study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which provides a structured, iterative methodology for data mining projects. The process comprises six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

3.1. Business Understanding

Objective: Develop an accurate and reliable fraud detection system for mobile money transactions that minimizes financial losses while maintaining acceptable false positive rates to avoid inconveniencing legitimate users.

3.2. Data Understanding

3.2.1. Dataset Overview

The dataset comprises 95,662 anonymized transactions from a mobile money platform in Kenya, collected over four months via automated logging. No personally identifiable information (PII) is included, ensuring compliance with data protection regulations.

3.2.2. Feature Description

The dataset contains 16 features spanning categorical, numerical, and temporal types:

Categorical: TransactionId, BatchId, AccountId, SubscriptionId, CustomerId, CurrencyCode, ProviderId, ProductId, ProductCategory, ChannelId

Numerical: Amount, Value, PricingStrategy

Temporal: TransactionStartTime

Target: FraudResult (binary: 1 = Fraud, 0 = Legitimate)

3.2.3. Class Distribution

Severe class imbalance is observed: Legitimate transactions: 95,469 (99.798Fraudulent transactions: 193 (0.202

This results in a fraud-to-legitimate ratio of approximately 1:495 (0.2% fraud). This severe skew presents a significant challenge for standard empirical risk minimization, as models can achieve 99.8% accuracy by trivially predicting the majority class (legitimate) while failing to detect any fraud attempts.

3.3. Data Preparation

3.3.1. Data Quality Assessment

No missing values or duplicate transactions were found. Temporal features were extracted from `TransactionStartTime`, including hour, day, month, weekday, and date components.

3.3.2. Outlier Detection and Treatment

IQR Outlier Detection Algorithm

Step 1: Calculate quartiles:
$Q1 = 25^{\text{th}}$ percentile, $Q3 = 75^{\text{th}}$ percentile

Step 2: Compute Interquartile Range:
$IQR = Q3 - Q1$

Step 3: Determine outlier boundaries:
Lower bound $= Q1 - 1.5 \times IQR$
Upper bound $= Q3 + 1.5 \times IQR$

Step 4: Identify outliers:
Data point x is an outlier if $x <$ Lower bound OR $x >$ Upper bound

Outliers in `Amount` and `Value` were identified using the Interquartile Range (IQR) method and retained as potential fraud indicators. Subsequently, robust scaling was applied to mitigate their influence.

3.3.3. Feature Engineering

Created temporal features from `TransactionStartTime` to capture time-based fraud patterns: Hour of day (0-23), Day of month, Month of year, Day of week (1-7), and Date component. These features enabled the model to learn temporal fraud patterns identified in exploratory analysis.

3.4. Exploratory Data Analysis

3.4.1. Temporal Fraud Patterns

Analysis revealed distinct temporal patterns in fraud occurrence, with higher rates during certain hours and days. Fraudulent transactions also exhibited different amount distributions compared to legitimate ones. Product categories and channels showed varying fraud susceptibility, informing subsequent feature selection.

3.4.2. Transaction Amount Analysis

We compared distributions of `Amount` and `Value` between fraud and legitimate transactions: Fraud transactions show distinct amount distributions. Certain ranges of amounts are associated with a higher propensity for fraud. Statistical differences validate `Amount` and `Value` as predictive features.

3.5. Data Preprocessing for Modeling

3.5.1. Feature Selection

Features were selected based on exploratory analysis and domain knowledge. Identifiers were excluded to prevent overfitting. Categorical variables were one-hot encoded, and numerical features were standardized using StandardScaler to ensure zero mean and unit variance.

3.5.2. Train-Test Split

The data were split into training (80%) and testing (20%) sets using stratified sampling to preserve class proportions. The training set was further used for cross-validation.

3.5.3. Feature Scaling

We applied **StandardScaler** to all numerical features to standardize them to zero mean and unit variance. For each feature value x with mean μ and standard deviation σ, the standardized value z is computed as:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

This transformation prevents features with larger scales from dominating model learning and is particularly important for algorithms sensitive to feature scales, such as Logistic Regression and Support Vector Machines.

3.5.4. Handling Class Imbalance

Given the severe class imbalance (1:495 fraud-to-legitimate ratio), we implemented a **class weighting** approach as the primary imbalance mitigation strategy. This technique assigns higher misclassification costs to the minority (fraud) class, forcing the algorithm to prioritize fraud detection during training. For class i, the weight w_i is computed as:

$$w_j = \frac{N}{k \times n_j} \tag{2}$$

where N is the total number of samples, k is the number of classes (2), and n_j is the number of samples in class j. For our minority fraud class, this yields a significantly higher weight ($w_{fraud} \approx 247$) compared to the majority class ($w_{legit} \approx 0.5$), effectively balancing the loss contribution.

We chose class weighting over synthetic oversampling techniques (e.g., SMOTE) because: (1) with a 1:495 ratio, SMOTE would require generating approximately 494 synthetic samples per real fraud case, risking unrealistic fraud patterns; (2) class weighting preserves data integrity by adjusting the learning algorithm rather than modifying the data; and (3) tree-based ensemble methods (Random Forest, XGBoost, LightGBM) handle class weights exceptionally well through their weighted loss functions.

As a complementary strategy, we performed **threshold optimization**, adjusting the classification threshold from the default 0.5 based on precision-recall trade-offs to align with business requirements for fraud detection rates versus false positive tolerance.

3.6. Model Development

3.6.1. Algorithm Selection

Five algorithms were selected for their complementary characteristics:

1. Logistic Regression (Baseline): A linear model that estimates the probability of fraud as:

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}^T \mathbf{x})}} \tag{3}$$

where $\mathbf{x}$ is the feature vector, β_0 is the intercept, and $\boldsymbol{\beta}$ represents feature coefficients. This model provides an interpretable baseline with coefficients indicating feature importance and direction, while remaining computationally efficient for real-time scoring.

2. Random Forest: An ensemble method that aggregates predictions from M decision trees:

$$\hat{y} = \frac{1}{M} \sum_{m=1}^{M} h_m(\mathbf{x}) \tag{4}$$

where $h_m(\mathbf{x})$ is the prediction of the m-th tree. Random Forest handles non-linear relationships and feature interactions through bootstrap aggregating (bagging), providing robustness to overfitting and feature importance rankings.

3. XGBoost and LightGBM (Gradient Boosting Machines): Advanced gradient boosting frameworks that minimize a regularized objective function:

$$\mathcal{L} = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

where ℓ is the loss function (weighted for class imbalance), $\hat{y}_i$ is the predicted value, and $\Omega(f_k)$ is the regularization term penalizing model complexity. Both frameworks excel at handling class imbalance through weighted loss functions and achieve superior performance through iterative boosting and regularization.

4. Linear Support Vector Machine (Linear SVM): A maximum-margin linear classifier that seeks to separate classes by optimizing a hinge-loss–based objective with regularization:

$$\min_{\mathbf{w},b} \left(\frac{1}{2}||\mathbf{w}||^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b)) \right) \tag{6}$$

where $\mathbf{w}$ is the weight vector defining the separating hyperplane, b is the bias term, $y_i \in \{-1, +1\}$ are the class labels, and C is a regularization parameter. Class imbalance is addressed through class-weighted penalty terms ($C_{fraud} > C_{legit}$), assigning higher misclassification costs to minority-class observations.

3.6.2. Model Training

We trained each algorithm on the class-weighted training data using **Stratified 5-Fold Cross-Validation** to ensure stable performance estimates. Hyperparameters were tuned using Grid Search within defined ranges. For Logistic Regression, we optimized the inverse regularization strength (C). Random Forest was configured with class weighting, 100 estimators, and max depth tuned between 10-50. For XGBoost and LightGBM, we utilized the `scale_pos_weight` parameter (set to $\approx$ 495) to explicitly handle class imbalance, while optimizing learning rates (0.01-0.1) and tree depth to prevent overfitting on the minority class.

3.7. Model Evaluation

3.7.1. Evaluation Metrics

We employed metrics specifically suited for imbalanced classification problems. Model predictions were evaluated using the confusion matrix, which categorizes predictions into True Positives (TP: correctly identified fraud), True Negatives (TN: correctly identified legitimate transactions), False Positives (FP: legitimate transactions flagged as fraud), and False Negatives (FN: missed fraud cases).

Given the severe imbalance, standard accuracy is misleading. Thus, the following metrics were used: **Precision** measures the accuracy of fraud predictions:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

Recall (or sensitivity) measures the percentage of actual fraud cases detected:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

The **F1-Score** provides the harmonic mean of precision and recall, balancing both metrics:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

We also evaluated **AUC-ROC** (Area Under the Receiver Operating Characteristic curve), which plots the true positive rate against the false positive rate across classification thresholds. Crucially for imbalanced datasets, we computed **PR-AUC** (Area Under the Precision-Recall curve), which focuses on minority class performance and is more informative than ROC-AUC when classes are severely imbalanced.

Finally, we considered cost-sensitive metrics reflecting the business impact of false negatives (missed fraud leading to financial losses) versus false positives (legitimate customers subjected to additional verification, causing friction).

3.7.2. Cross-Validation

Employed stratified k-fold cross-validation to ensure robust performance estimates while maintaining class proportions in each fold.

3.8. Threshold Optimization

The classification threshold was tuned by analyzing precision-recall curves to balance fraud detection (recall) against false positives (precision).

3.9. Deployment Considerations

Discussed practical aspects for production deployment:

- **Real-time Scoring**: Low-latency prediction requirements
- **Model Monitoring**: Tracking performance degradation and fraud pattern drift
- **Periodic Retraining**: Updating models with new fraud patterns
- **Explainability**: Providing justification for fraud alerts
- **Integration**: API design for system integration

3.10. Deployment Prototype: Streamlit Application

To validate the practical feasibility of the proposed fraud detection framework, we developed an interactive web application using Streamlit. This prototype serves as a proof-of-concept for real-time fraud scoring and analyst review.

Key Features:

- **Real-time Inference**: Loads the trained LightGBM model and scaler artifacts to generate fraud probability scores for new transactions instantly.
- **Feature Consistency**: Implements an identical feature engineering pipeline to the training phase, ensuring consistent transformation of raw transaction data (e.g., temporal features, amount ratios) during inference.
- **Analyst Interface**: Provides a dashboard for visualizing transaction risk, including a "Fraud Probability Gauge" and key risk factors, enabling fraud analysts to make informed decisions.
- **Batch Processing**: Supports bulk upload of transaction logs for retrospective analysis, mimicking the batch processing workflows used in production environments.

This application demonstrates that the complex feature engineering and model inference steps can be executed with low latency, supporting the "Real-time Scoring" requirement outlined in the deployment considerations.

References

Afriyie, J.K., Tawiah, K., Prah, A.K., Annan, E., Weber, F., 2023. A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. Decision Analytics Journal 6, 100163. doi:10.1016/j.dajour.2023.100163.

Ahmed, K.H., Axelsson, S., Li, Y., Sagheer, A.M., 2025. A credit card fraud detection approach based on ensemble machine learning classifier with hybrid data sampling. Machine Learning With Applications 20, 100675. URL: https://doi.org/10.1016/j.mlwa.2025.100675, doi:10.1016/j.mlwa.2025.100675.

Al-Asadi, M., Alissa, A.E., Bhushan, B., Al-Azzawi, M., 2025. Enhancing financial fraud detection using xgboost and advanced data balancing techniques, in: 2025 1st International Conference on Secure IoT, Assured and Trusted Computing (SATC), IEEE. pp. 1–16. doi:10.1109/SATC65530.2025.11137062.

AL-Dahasi, E.M., Alsheikh, R.K., Khan, F.A., Jeon, G., 2024. Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation. Expert Systems doi:10.1111/exsy.13682.

Al-Khasawneh, M.A., Faheem, M., Alsekait, D.M., Abubakar, A., Issa, G.F., 2025. Hybrid neural network methods for the detection of credit card fraud. Security and Privacy 8, e500. doi:10.1002/spy2.500.

Albalawi, T., Dardouri, S., 2025. Enhancing credit card fraud detection using traditional and deep learning models with class imbalance mitigation. Frontiers in Artificial Intelligence 8, 1643292. URL: https://doi.org/10.3389/frai.2025.1643292, doi:10.3389/frai.2025.1643292.

Amazon Web Services, 2024. Implementing Real-Time Fraud Detection on AWS. Whitepaper. Amazon Web Services. URL: https://docs.aws.amazon.com/whitepapers/latest/real-time-fraud-detection-on-aws/real-time-fraud-detection-on-aws.pdf.

Arzu, F., Bajwa, M., Obaidullah, Waheed, A., Alam, F., Ali, M., Khan, A., 2025. Real-time financial fraud detection: An intelligent data-driven framework integrating machine learning, stream processing, and big data analytics for high-velocity transaction monitoring. The Asian Bulletin of Big Data Management 5, 124–154. doi:10.62019/f3c0g313.

Azamuke, D., Katarahweire, M., Bainomugisha, E., 2025. Financial fraud detection using rich mobile money transaction datasets, in: E-Infrastructure and e-Services for Developing Countries (AFRICOMM 2023), Springer. pp. 234–248. doi:10.1007/978-3-031-81573-7_16.

Azamuke, D., Katarahweire, M., Businge, J.M., Kizza, S., Opio, C., Bainomugisha, E., 2023. Refining detection mechanism of mobile money fraud using momtsim platform. doi:10.1007/978-3-031-57639-3_3.

Bonde, L., Bichanga, A.K., 2025. Improving credit card fraud detection with ensemble deep learning-based models: A hybrid approach using smote-enn. Preprints doi:10.20944/preprints202501.0234.v1.

Botchey, F.E., Qin, Z., Hughes-Lartey, K., 2020. Mobile money fraud prediction—a cross-case analysis on the efficiency of support vector machines, gradient boosted decision trees, and naïve bayes algorithms. Information 11. doi:10.3390/INFO11080383.

Botchey, F.E., Qin, Z., Hughes-Lartey, K., Ampomah, E.K., 2022. Predicting fraud in mobile money transactions using machine learning: The effects of sampling techniques on the imbalanced dataset. Informatica 45. doi:10.31449/inf.v45i7.3179.

Chauhan, Y., Uppal, M., Gupta, D., Mapari, S., Saini, S., 2025. Enhancing accuracy of financial fraud detection in mobile transactions using xgboost algorithm, in: 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI), pp. 302–307. doi:10.1109/IC3ECSBHI63591.2025.10991168.

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

Chen, Z., Duan, J., Kang, L., Qiu, G., 2022. Class-imbalanced deep learning via a class-balanced ensemble. IEEE Transactions on Neural Networks and Learning Systems 33, 5626–5640. doi:10.1109/TNNLS.2021.3071122.

Cheng, D., Zou, Y., Xiang, S., et al., 2025. Graph neural networks for financial fraud detection: a review. Frontiers of Computer Science 19, 199609. doi:10.1007/s11704-024-40474-y.

Cochrane, N., Gomez, T., Warmerdam, J., Flores, M., McCullough, P., Weinberger, V., Pirouz, M., 2021. Pattern analysis for transaction fraud detection. doi:10.1109/CCWC51732.2021.9376045.

Dhasaratham, M., Balassem, Z.A., Bobba, J., Ayyadurai, R., Sundaram, S.M., 2024. Attention based isolation forest integrated ensemble machine learning algorithm for financial fraud detection. doi:10.1109/iacis61494.2024.10721649.

Dinesh, M.V., 2024. Comparative analysis of machine learning models for credit card fraud detection. International Journal of Engineering Research in Technology 13.

Doddamani, S.S., Girish, K.K., Bhowmik, B., 2024. Money laundering detection in imbalanced e-wallet transactions with threshold optimization. doi:10.1109/i2ct61223.2024.10544197.

Emran, A.M., Rubel, M.T.H., 2024. Big data analytics and ai-driven solutions for financial fraud detection: Techniques, applications, and challenges doi:10.70937/faet.v1i01.40.

Green, A., 2025. Ai-driven financial intelligence systems: A new era of risk detection and strategic analysis. doi:10.31219/osf.io/ynph2_v1.

GSMA, 2024. State of the Industry Report on Mobile Money 2024. www.gsma.com. Accessed: 2025-12-15.

Gupta, A., 2025. Real-time online payment fraud detection using machine learning algorithms in financial systems. Indian Scientific Journal of Research in Engineering and Management 9, 1–9. doi:10.55041/ijsrem52560.

Hanbali, N., El-Yahyaoui, A., 2025. Advanced machine learning and deep learning approaches for fraud detection in mobile money transactions. Innovations in Systems and Software Engineering 21, 333–353. URL: https://doi.org/10.1007/s11334-025-00605-5, doi:10.1007/s11334-025-00605-5.

Hájek, P., Abedin, M.Z., Sivarajah, U., 2022. Fraud detection in mobile payment systems using an xgboost-based framework. Information Systems Frontiers , 1–19doi:10.1007/s10796-022-10346-6.

(IMF), I.M.F., . Battling cybercrime: A global fight. finance development .

Jeyachandran, P., Akisetty, A.S.V.V., Subramani, P., Goel, O., Singh, D.S.P., Shrivastav, E.A., 2024. Leveraging machine learning for real-time fraud detection in digital payments. Integrated Journal for Research in Arts and Humanities 4, 70–94. doi:10.55544/ijrah.4.6.10.

Kandi, K., 2025. Enhancing performance of credit card model by utilizing lstm networks and xgboost algorithms. Machine Learning and Knowledge Extraction 7, 20. doi:10.3390/make7010020.

Kaur, G., 2025. Enhancing fraud detection in portable wallet payment systems using machine learning: A hybrid approach. Deleted Journal , 1317–1330doi:10.52783/cana.v32.5156.

Kodete, C.S., 2023. Mathematical modelling of fraud detection in mobile financial transactions using deep learning. International Journal of Scientific Research in Science and Technology , 724–739doi:10.32628/ijsrst2302524.

Kumar, S., 2024. Enhanced fraud detection in financial transactions using hyperparameter-tuned random forests. doi:10.1109/icccnt61001.2024.10725958.

Liu, C., Tang, H., Yang, Z., Zhou, K., Cha, S., 2025. Big data-driven fraud detection using machine learning and real-time stream processing. doi:10.48550/arXiv.2506.02008.

Lokanan, M.E., 2023. Predicting mobile money transaction fraud using machine learning algorithms. Applied AI Letters 4, e85. doi:10.1002/ail2.85.

Lu, J., 2024. Improving fraud detection in mobile payments with machine learning ensembles, in: 2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS), pp. 851–854. doi:10.1109/EIECS63941.2024.10800126.

Mambina, I.S., Ndibwile, J.D., Michael, K.F., 2022. Classifying swahili smishing attacks for mobile money users: A machine-learning approach. IEEE Access 10, 83061–83074. doi:10.1109/ACCESS.2022.3196464.

McKinsey & Company, 2023. The future of real-time payments: Fighting fraud in a faster world. McKinsey on Payments URL: https://www.mckinsey.com/industries/financial-services/our-insights/the-future-of-real-time-payments-fighting-fraud-in-a-faster-world.

Mollik, E., Majeed, F.A., 2025. Ai-driven cybersecurity in mobile financial services: Enhancing fraud detection and privacy in emerging markets. Journal of Cybersecurity and Privacy 5, 77. doi:10.3390/jcp5030077.

Muqattash, R., Kharbat, F., 2023. Detecting mobile payment fraud: Leveraging machine learning for rapid analysis, in: 2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 1–5. doi:10.1109/SNAMS60348.2023.10375448.

National Kenya Computer Incident Response Team Coordination Centre (National KE-CIRT/CC), 2023. Kenya cybersecurity report 2023. URL: https://www.ca.go.ke/wp-content/uploads/2024/03/Cybersecurity-Report-2023.pdf. retrieved from the Communications Authority of Kenya website.

Neza, F., Joseph, A., Joseph, M., 2022. E-money security dilemma: Advanced cybersecurity mechanisms and legacy mobile payments in sub-saharan africa, in: Proceedings of the International Conference on Web Intelligence (ICWI 2022). doi:10.33965/ac_icwi2022_2022081013.

Papasavva, A., Lundrigan, S., Lowther, E., et al., 2025. Applications of ai-based models for online fraud detection and analysis. Crime Science 14. doi:10.1186/s40163-025-00248-8.

Psychoula, I., Gutmann, A., Mainali, P., Lee, S., Dunphy, P., Petitcolas, F., 2021. Explainable machine learning for fraud detection. Computer 54. doi:10.1109/MC.2021.3081249.

Ramesh, A., K, M.R., 2025. A stacked lightgbm-xgboost model with shap-based fraud detection for financial transactions. International Scientific Journal of Engineering and Management 4, 1–7. doi:10.55041/isjem02496.

Raturi, A., 2024. A comparative analysis of machine learning algorithms for credit card fraud detection, in: 2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC), IEEE. pp. 1–6. doi:10.1109/ICECCC61767.2024.10593936.

Renukadevi, S., Manujakshi, B.C., Shashidhar, T.M., Sivakumar, N., 2025. Fraud detection in financial transactions using gradient boost with hybrid optimization. Journal of Machine and Computing , 2328–2344doi:10.53759/7669/jmc202505181.

Sa'adah, S., Pratiwi, M.S., 2020. Classification of customer actions on digital money transactions on paysim mobile money simulator using probabilistic neural network (pnn) algorithm. doi:10.1109/ISRITI51436.2020.9315344.

Sanni, M.L., Akinyemi, B.O., Olalere, D.A., Olajubu, E.A., Aderounmu, G.A., 2023. A predictive cyber threat model for mobile money services. Annals of Emerging Technologies in Computing 7, 40–60. doi:10.33166/aetic.2023.01.004.

Sariat, A.F., Siddique, I.J., Hossain, M., Islam, M.M., Rahman, T., 2025. Ai driven fraud detection in financial ecosystems: A hybrid machine learning framework, in: 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1–8. doi:10.1109/ECCE64574.2025.11013808.

Shinde, Y., Chadha, A.S., Shitole, A., 2021. Detecting fraudulent transactions using hybrid fusion techniques.

Silva, J.C.S., Macêdo, D., Zanchettin, C., Oliveira, A.L.I., Filho, A.T.D.A., 2021. Multi-class mobile money service financial fraud detection by integrating supervised learning with adversarial autoencoders. doi:10.1109/IJCNN52387.2021.9533313.

Souran, M., Shah, R.S., 2025. Fraud detection in mobile payments using a hybrid machine learning model, in: 2025 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), pp. 1–6. doi:10.1109/AMATHE65477.2025.11081192.

Sundaravadivel, P., Isaac, R.A., Elangovan, D., KrishnaRaj, D., Rahul, V.V., Raja, R., 2025. Optimizing credit card fraud detection with random forests and SMOTE. Scientific Reports 15, 17851. URL: https://doi.org/10.1038/s41598-025-00873-y, doi:10.1038/s41598-025-00873-y.

Tagbo, S.K., Adekoya, A.F., Mensah, P.K., 2024. Mitigating mobile money social engineering attacks using machine learning. doi:10.22541/au.171929692.21699330/v1.

Theodorakopoulos, L., Theodoropoulou, A., Tsimakis, A., Halkiopoulos, C., 2025. Big data-driven distributed machine learning for scalable credit card fraud detection using pyspark, xgboost, and catboost. Electronics 14, 1754. doi:10.3390/electronics14091754.

UK Finance, 2024. Annual fraud report 2024. URL: https://www.ukfinance.org.uk/system/files/2024-05/Annual%20Fraud%20Report%202024.pdf.

Wang, J., Yang, C., 2022. Financial fraud detection based on ensemble machine learning, in: 2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), pp. 1–6. doi:10.1109/DASC/PiCom/CBDCom/Cy55231.2022.9928001.

Wold, B.J., 2023. Fraud detection in mobile banking based on artificial intelligence. doi:10.1007/978-3-031-35314-7_48.

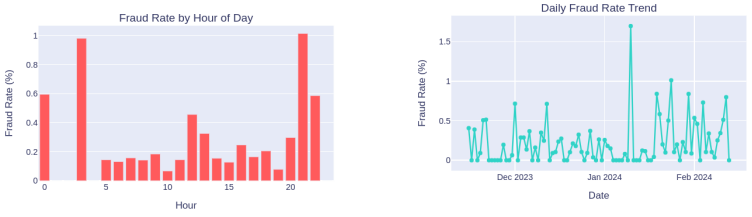
Yussif, N., Takyi, K., Gyening, R.O.M., Boadu-Acheampong, S.I., 2025. Advanced mobile money fraud detection using cnn-bilstm and optimized sgd with momentum. AJIT-e: Online Academic Journal of Information Technology 16, 207–231. doi:10.5824/ajite.2025.03.002.x.

Zhao, X., Liu, Y., Zhao, Q., 2024a. Improved lightgbm for extremely imbalanced data and application to credit card fraud detection. IEEE Access 12, 159316–159335. doi:10.1109/ACCESS.2024.3487212.

Zhao, X., Zhang, Q., Zhang, C., 2024b. Enhancing transaction fraud detection with a hybrid machine learning model. doi:10.1109/icetci61221.2024.10594463.

Zheng, Q., Chen, Y., Cao, J., Xu, Y., Xing, Q., Jin, Y., 2024a. Advanced payment security system: Xgboost, catboost and smote integrated. doi:10.48550/arxiv.2406.04658.

Zheng, Q., Yu, C., Cao, J., Xu, Y., Xing, Q., Jin, Y., 2024b. Advanced payment security system: Xgboost, lightgbm and smote integrated. doi:10.1109/metacom62920.2024.00063.



(a) Hourly Trend (b) Daily Trend

Figure 1: Temporal analysis of fraud occurrence.