Applying Machine Learning to Enhance Fraud Detection in Kenyan Digital Banking

Michael Imende Edward
169405

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Data Science and Analytics at Strathmore University

Institute of Mathematical Sciences and iLab Africa
Strathmore University
Nairobi, Kenya

June, 2025

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other university. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Student's Name: **Michael Imende Edward**

Sign: _____ Date: 27/05/2025

Approval

The dissertation of **Michael Imende Edward** was reviewed and approved by the following:

Dr. Kennedy Senagi,
Lecturer, Institute of Mathematical Sciences,
Strathmore University.

Dr. Godfrey Madigu,
Dean, Institute of Mathematical Sciences,
Strathmore University.

Prof. Benard Shibwabo,
Director of Graduate Studies,
Strathmore University.

Abstract

In Kenya, leading financial institutions have lost millions due to financial fraud. Financial fraud occurs when someone (for example, a client) loses their money or assets through deception. Despite the numerous benefits, electronic transactions have created space for malicious actors to take advantage of questionable security features to get away with financial fraud. Conventional techniques cannot address the challenges such transactions present. They are slow, costly, and inaccurate, making them unreliable in this new space. Machine Learning (ML) techniques offer hope regarding preventing these crimes from growing and wreaking havoc in the industry. They are fast, accurate, and can adapt through learning to prevent new crimes. This study investigated past fraudulent financial transactions in the Kenyan finance market, identified the attributes and features contributing to fraudulent financial transactions, and developed a reasonable approach that relies on ML techniques to detect fraudulent transactions early. The research evaluated algorithms that could assist in detecting and classifying transactions accurately, relying on datasets from the Kenya mobile banking sector. The Synthetic Minority Oversampling Technique (SMOTE) technique was used to address the data imbalance within this dataset. The dataset was split into training and test data, with feature extraction ensuring that this division was accurate and precise. Several algorithms were explored, and their performance was assessed. Before hyperparameter tuning, Random Forest achieved an Area Under the Curve (AUC) of 0.995 but failed to detect any fraudulent transactions (precision and recall for class 1 were 0.32 and 0.29, respectively), resulting in a macro F1-score of 0.63, while Logistic Regression reached a precision of 0.12 and recall of 0.65 for fraud with an overall accuracy of 73% and a macro F1-score of 0.52. After tuning, the optimized XGBoost model achieved an overall accuracy of 83%, with a fraud precision of 0.81, a recall of 0.86, and an F1-score of 0.83 for the minority class. In addition, XGBoost's macro average F1-score improved to 0.83, and its log-loss decreased to 0.185, indicating better stability and balanced performance across classes. Adjusting the decision threshold further enhanced fraud detection, increasing recall to 0.95 with a corresponding precision of 0.74 and an F1-score of 0.83. Overall, these performance numbers confirm that XGBoost is the best-performing model for detecting fraudulent

transactions in the study. Logistic regression was used to predict the outcome of events; random forest combined multiple decision trees to achieve a single result; Artificial Neural Network (ANN) assisted in recognizing patterns and solving common problems; The results showed that the algorithms efficiently and accurately detected financial fraud. Model selection was followed by training, model performance evaluation, and model tuning and optimization to enhance generalization ability. The model was validated by feeding it with actual transactions and assessing its efficacy in flagging fraud and non-fraud activities. The model was deployed behind a mobile and web application displaying the model evaluation results.

Keywords: Cross-Industry Standard Process for Data Mining (CRISP-DM), XGBoost, SMOTE, ANN, Random Forest

Table of Contents

List of Abbreviations

ADASYN Adaptive Synthetic Sampling

AI Artificial Intelligence

ANN Artificial Neural Network

ANNs Artificial Neural Networks

AUC Area Under the Curve

BCR Balanced Classification Rate

CNNs Convolutional Neural Networks

CRISP-DM Cross-Industry Standard Process for Data Mining

DBSCAN Density-Based Spatial Clustering of Applications with Noise

EDA Exploratory Data Analysis

GBMs Gradient Boosting Machines

Kshs Kenya Shillings

ML Machine Learning

RAMO Ranking Minority Over-sampling Boost

RNNs Recurrent Neural Networks

ROC Receiver Operating Characteristic

ROI Return on Investment

SIM Subscriber Identity Module

SMOTE Synthetic Minority Oversampling Technique

SVM Support Vector Machine

SVMs Support Vector Machines

Chapter 1: Introduction

1.1 Background

The fast growth in digital finance technology has greatly impacted the sector as financial transactions become easier and more accessible. Mobile money platforms are a great case in point as of the transformational rise in digital financial transactions, with Kenya leading in mobile banking globally (Botchey et al., 2020; Lokanan, 2023). With Kenya's mobile money penetration exceeding 83% and circulating billions annually, financial fraud remains a threat. A prime example is the recent losses of Kenya Shillings (Kshs) 1.6 billion by Equity and ABSA banks through fraud (Muiruri, 2023; Botchey et al., 2020). Hackers exploit gaps in electronic networks, leading to massive financial losses, erosion of consumer confidence, and reputation loss in financial institutions(Ngai et al., 2011; Albashrawi, 2016).

Advancements in telecommunications continue to pave the way for adjusting and improving everyday living. In the financial sector, telecommunications has enabled customers to access their accounts remotely, easing transactions and boosting satisfaction. However, cybercrimes targeting financial systems have seen a surge in fraudulent transactions targeting financial institutions and other industry players. Electronic transactions have allowed criminals to use questionable security features to get away with financial fraud (Ryman-Tubb et al., 2018). Fraudulent financial activities can cause financial institutions to lose money and relevance in the market. The advent of digital banking has exacerbated this issue, making it imperative to develop sophisticated systems. Traditional systems are often insufficient due to their inherent capacity to address evolving tactics.

According to Hilal et al., financial fraud is an intentional act that betrays the trust of stakeholders and causes financial harm (Hilal et al., 2022). Financial fraud is not limited to financial institutions. Instead, it is a practice that is rife in all sectors of the economy. Cases of financial fraud have been increasingly exponential in recent years (Choi and Lee, 2018). Efforts to slow this growth have not yielded the expected results (Ryman-Tubb et al., 2018). Its persistence has adversely impacted the world in terms of the money lost in avoidable ways. Many fraud detection methods have risen to address this problem (Hilal et al., 2022). Most of these methods have

been underwhelming, a reality demonstrated by the rising cases of financial fraud (Al-Hashedi and Magalingam, 2021). Studies examining this trend have also increased, but they, too, have been inefficient (Lokanan, 2023).

Fraud affects every business segment worldwide, from banks to insurance providers, tax agencies, and corporate organizations. Financial fraud is generally complex and sophisticated, making it too complex for traditional detection means, such as manual audits and rule-based systems (Albashrawi, 2016). The primary reason is that conventional detection tools are marred by raw inefficiencies, inaccuracies, and the inability to deal with rising fraud sophistication today. These realities justify the need for modern and effective detection tools and strategies that match the sophistication of modern financial crimes (Ngai et al., 2011).

Leading financial institutions in Kenya have lost millions due to fraudulent financial transactions. Notable Kenyan financial institutions that have been a target of fraudulent financial transactions in the past few years include ABSA Bank and Equity Bank. In 2013, the Nation reported that rogue bank employees stole over Kshs. 1.5 billion from financial institutions between 2012 and 2013 (Nation, 2013). Additionally, the Banking Fraud Investigations Department claimed that between the same periods, fraudsters stole over Kshs. 1.49 billion from customer accounts. In 2022, ABSA Bank lost more than Kshs. 107.7 million to fraudsters, although the institution managed to recover Kshs. 59.9 million (Muiruri, 2023). Equity Bank Kenya also reported losses of Kshs. 271.22 million due to fraud involving hacking into card data, with criminals primarily using stolen card details to defraud unsuspecting victims and shop online (Munyua, 2013; Delamaire et al., 2009). The rising number of fraud cases targeting Kenyan banks demonstrates the essence of the issue to the country's economy. As cybercriminals use electronic transactions to steal from institutions, ML algorithms offer a sustainable solution to detecting and preventing these crimes.

In Kenya, the necessity to enhance fraud detection controls is accompanied by the increased reliance on mobile money platforms and heightened sophistication of cybercrimes. The rapid adoption of digital financial services, particularly mobile money, has created new avenues for fraudsters to exploit vulnerabilities in the system. As more Kenyans embrace mobile banking for its convenience and accessibility, the potential for fraudulent activities has grown exponentially. Consequently, financial institutions

face the pressing need to implement robust fraud detection mechanisms that can keep pace with the evolving tactics employed by cybercriminals. The rising cybercrimes targeting Kenya's financial institutions testify to the exigency for advanced technical controls in the country's digital financial ecosystem.

Fraudulent cases in the financial sector have increased, particularly those involving stolen bank details and swapped Subscriber Identity Module (SIM) cards. This trend has prompted a rethinking of how to prevent this vice from becoming rampant. Machine learning has surfaced as an efficient solution to the ongoing fraudulent activities in the sector (Chaquet-Ulldemolins et al., 2022; Da'u and Salim, 2020). There has been a sharp increase in the use of supervised and unsupervised ML to predict these activities before they surface (Choi and Lee, 2018; Zeng and Tang, 2021). Among these two broad methods, classification methods are the most commonly used in detecting fraud (Lokanan et al., 2019). Generally, a ML model goes through two processes before being ready. The first stage is where it is fed with data and trained on it. Testing is the subsequent step after training. Here, the model is tested on samples to determine its efficiency and accuracy in terms of classification (Ashtiani and Raahemi, 2021; Hilal et al., 2022). This study attempts to identify ML techniques used in fraud detection to explore gaps and discover research trends in this area.

Over time, there has been an upsurge in research on financial fraud detection. A trend easily observable in the literature is the large number of articles on this topic in recent times (Delamaire et al., 2009; Zhang and Zhou, 2004; Raj and Portia, 2011). This trend can be linked to the growing need to address this problem in the financial sector. Delamaire and colleagues (Delamaire et al., 2009) reviewed fraudulent activities on credit cards, examining the nature and impact of each. They conclude with proposals of how to address them. In their 2004 study on ML methods used to detect fraudulent transactions, Zhang and Zhou (Zhang and Zhou, 2004) examined their effectiveness and the areas they have been most effective. Raj and Portia (Raj and Portia, 2011) examined the credit fraud cases in the financial sector and the ML techniques that have been used to address these problems. Phua and colleagues investigated how ML techniques have been applied to different fraud cases (Phua et al., 2010). Their study demonstrates the increasing reliance on these approaches to enhance security. Various studies have shown that financial fraud has not been

limited to the financial sector. The health industry has witnessed a sharp growth in such occurrences (West and Bhattacharya, 2016). Abdallah et al. (Abdallah et al., 2016) examined the approaches used in the healthcare sector to identify and prevent fraudulent activities. They found that healthcare institutions are also rapidly embracing ML approaches. Popat and Chaundry (Popat and Chaudhary, 2018) review why people choose online payments, why fraudsters target such payments, the methods they use, and how ML methods can combat such fraud.

Research on ML classification techniques has enhanced the knowledge of their use cases and challenges. It has also increased understanding of their methodologies and functionality. The study by Ryman-Tubb et al. (Ryman-Tubb et al., 2018) examined various ML techniques used in detecting fraud. Their study on the latest ML methods focused specifically on transactional volumes. They found that while many methods exist, only a limited number (eight) are used in the sector. Albashrawi and Lowell (Albashrawi, 2016) reviewed multiple studies over ten years. The goal was to understand how financial fraud is detected in the industry. The researchers focused on data mining techniques. Despite the number of sources studied and the years covered, the findings are unreliable because many errors were made, including ignoring the positives and negatives of data mining.

The literature on financial fraud has focused chiefly on specific finance areas. Many areas have been left out or are scarcely covered in the existing body (Gyamfi and Abdulai, 2018). There are many ways to explain this trend. However, the most logical reason is the problem of imbalanced datasets, which make it challenging for studies to focus on particular crimes (Carneiro et al., 2015; Iyer et al., 2011; Patil et al., 2018). This crucial gap must be addressed to increase understanding of the issue and devise ways to address it. More recently, a study (Ashtiani and Raahemi, 2021) attempted to address this gap by taking a multidisciplinary approach to the problem. This study took a different approach from this. Unlike their research on financial statements, this one focused on diverse financial transactions.

Financial fraud is estimated to run into billions of dollars annually, and there is an urgent need for fresh solutions. Recent developments in Artificial Intelligence (AI) and ML offer real promise for fraud detection automation. Machine learning algorithms can identify fraud patterns typical to most frauds while examining large volumes of

data, thus becoming increasingly accurate and quicker to detect (Hilal et al., 2022). Machine learning algorithms can identify common fraud patterns by analyzing huge datasets and thus become more precise and efficient in their detection. Some leeway to counter challenges still exists, particularly when dealing with the skewed nature of the fraud-related data sets where legal transactions occur many times more often than fraud.

The current study addresses these enduring challenges by applying machine learning methods to detect fraudulent transactions within the Kenyan mobile money economy. The study explored variables relevant to fraudulent transactions and compared the performance of algorithms such as Random Forest, XGBoost, ANN, and Logistic Regression (Afriyie et al., 2023; Achary and Shelke, 2023). In addition, data imbalance issues were handled by techniques such as SMOTE that enable appropriate model performance (Phua et al., 2010; Popat and Chaudhary, 2018). Utilizing anonymized datasets from Kenyan financial institutions, this study demonstrated the creation of a predictive model with the capability to identify fraud in real time based on transaction frequency, location, and values. Aside from detection, the approach outlined in this paper is practice-oriented, combining machine learning models with web and mobile applications that facilitate real-time fraud detection. The applications are designed to give valuable insights that empower financial institutions to act in time against fraudulent attempts and strengthen the resilience of Kenya's financial infrastructure against fraud.

This research contributes to economic and social advancements by preventing financial losses, enhancing operational efficiency, and restoring confidence in electronic transactions (Ryman-Tubb et al., 2018). The findings demonstrate that the proposed model can be extended to other forms of financial fraud, such as credit card fraud and money laundering, by providing generalizable fraud detection solutions. Furthermore, this study supports global efforts to combat financial crime by offering a responsive and sustainable approach to securing digital financial systems (Albashrawi, 2016; Ngai et al., 2011).

1.2 Problem Statement

Financial fraud has increased and diversified globally, affecting all sectors critical to the world economy. Different methods have been employed to slow or reverse this trend, but the results of such efforts have been underwhelming. Current detection methods, such as manual audits and rule-based systems, have proven ineffective in dealing with the increasing sophistication and volume of fraudulent activities. These methods are often slow, error-prone, and unable to adapt to the evolving tactics employed by fraudsters, leading to substantial financial losses.

Machine learning has emerged as a promising approach to address the limitations of traditional fraud detection techniques. By leveraging advanced algorithms and large datasets, ML models can automatically identify suspicious patterns, adapt to new fraud schemes, and make real-time predictions to prevent financial losses. However, implementing ML for fraud detection comes with its own set of challenges, such as handling imbalanced datasets, ensuring model interpretability, and keeping up with the ever-changing fraud landscape.

The current study investigated advanced ML algorithms for detecting various types of financial fraud. It focused on improving detection accuracy in imbalanced datasets and examined the potential of unsupervised learning approaches like clustering. The research aimed to enhance fraud detection systems, lower monetary losses in the financial sector, and contribute to enhanced fraud prevention measures by studying real-world data and applying unique algorithms.

Existing studies on fraud in the financial sector rarely cover all types. They mostly focus on specific forms of fraud, such as credit card fraud, which limits the generalizability of their findings and the effectiveness of the proposed solutions. To address this gap, the current study takes a more comprehensive approach, investigating diverse types of financial fraud and developing machine learning models that can adapt to various fraudulent activities.

In the context of Kenya, where mobile transactions have become increasingly prevalent, the need for robust fraud detection mechanisms is particularly pressing. Cybercriminals have exploited technological weaknesses to defraud institutions, resulting in substantial financial losses. The current study focuses on leveraging ML algorithms to identify fraudulent patterns in Kenyan mobile transactions, with the aim

of reducing the frequency and impact of such activities. By utilizing anonymized data from Kenyan banks and incorporating real-time detection capabilities, the developed model seeks to provide a practical solution to the challenges faced by the country's financial sector.

1.3 Research Objectives

The main objective of the current study was to develop and evaluate advanced ML algorithms for detecting financial fraud in the Kenyan context, with a primary focus on supervised and ensemble methods. Specifically, the research aimed to:

(i) To investigate how ML has been applied in fraud detection.

(ii) To develop a comprehensive ML model for fraud detection that utilizes supervised learning, focusing on ensemble methods.

(iii) To deploy the fraud detection model on web and mobile platforms.

1.4 Research Questions

This study answered the following questions:

(i) How is ML currently applied in fraud detection, and which fraud types are most commonly addressed?

(ii) How can a comprehensive ML model for fraud detection be developed utilizing supervised learning with a focus on ensemble methods?

(iii) What are the challenges and best practices for deploying an ML fraud detection model in a mobile and web application environment?

1.5 Significance of the Research

This study's significance stems from its attention to the difficulties in effectively identifying and stopping financial fraud. The first challenge it addresses is imbalanced data. In the financial sector, as in the healthcare sector, genuine transactions outnumber fraudulent ones. This reality results in a situation where the datasets available are

insufficient for accuracy and efficiency. The research focuses on strategically addressing this real-world challenge to enhance fraud detection. It aims to improve the models' ability to detect suspicious patterns.

Second, this research examines the effectiveness of unsupervised learning strategies. Their underutilization in fraud detection is inexplicable, given their efficiency in producing positive results. The advantage of these techniques is that they require no human intervention. This ability to discover hidden patterns in information (similarities and differences) makes them ideal for data analysis.

Finally, this research proposes a hybrid approach, combining the strengths of both methods. The rationale behind this approach is that integrating their strengths enhances system effectiveness. The research benefits the financial sector by reducing financial losses, improving operational efficiency, and helping build a system that protects banks. With machine learning, it becomes easier to identify gaps in the system and predict potential future vulnerabilities. Many players in the finance industry have implemented machine learning algorithms to address fraudulent cases related to insurance fraud, credit card fraud, telecom subscription fraud, and other fraudulent financial transactions (Ali et al., 2022). As fraudulent financial activities continue to rise in Kenya, using ML algorithms may provide a solution to mitigate the financial burden imposed on financial institutions. Recently, the high levels of fraudulent financial activities involving major financial organizations in Kenya, such as ABSA Bank and Equity Bank, demonstrate the need for improved security measures to address the issue. The current research identifies a reasonable basis for using ML algorithms to stop fraudulent transactions. However, the efficiency of fraud detection algorithms depends on their ability to detect fraud in real time rather than merely identifying past occurrences.

1.6 Research Hypothesis

The primary hypothesis is that ML techniques are more efficient in fraud detection than traditional methods. The advantages of ML techniques become evident when dealing with imbalanced data and when unsupervised learning is required to generate accurate results. Regardless of its nature and complexity, financial fraud cannot be efficiently addressed by manual methods or rule-based systems. These conventional

methods lack reliability in handling massive datasets and identifying emerging fraud patterns. In cases where fraud is detected, the results are often irrelevant due to delayed detection.

The research posits that using advanced machine learning algorithms enhances fraud detection. Notable improvements are expected in accuracy (the frequency of predicting correct outcomes) and speed (the time taken to detect fraud). Using supervised and unsupervised techniques increases the expected efficiency. The research also hypothesizes that combining these methods (supervised and unsupervised) reduces the likelihood of missing unusual patterns in transactions.

The research also explored how machine learning addresses imbalanced data. The chief issue with this data was that it increased the odds of biased outcomes, impacting the reliability and efficacy of the ML models. The research suggested that using techniques such as SMOTE could address this imbalance, which is common in supervised machine learning. SMOTE reduced false positives while making identifying and trapping fraud easier. Generally, the research demonstrated that ML approaches are more efficient than conventional methods in fraud detection. The industry requires models that detect fraud when data is at rest, in transit, and in real time; machine learning models fulfill this need.

1.7 Scope and Limitations

The research examined how ML algorithms detect financial fraud in varied domains. The research was limited to the most common frauds, meaning it did not consider all crimes. The most common frauds included credit card fraud, insurance fraud, and money laundering. The research focused on the accuracy and efficiency of detecting such frauds using ML techniques. The research explored supervised and unsupervised techniques and how they could be applied to address fraud in this context. It also examined how to deal with imbalanced data in classification. This problem was primarily caused by the rarity of the event being studied. In the financial sector, legitimate transactions outnumbered fraudulent ones, making it challenging to learn to identify the minority class.

This research had four notable limitations. First, it relied on existing datasets to develop efficient models. The problem with existing datasets was that they did not

cover all types of fraud. This reality resulted in the model being forced to generate educated guesses, which could be either wrong or right. Second, the research focused only on cases with sufficient datasets. Working with such cases made it easier to improve model performance because the model had many examples to learn from. It also reduced the chances of underfitting, a phenomenon where the model struggled to identify essential links in the data.

While this study utilized anonymized data from a well-regarded financial institution, it is essential to critically examine the representativeness of the dataset and the ethical implications of using such data for real-time fraud detection. The anonymization process, while necessary for protecting consumer privacy, may inadvertently introduce biases or limitations in the dataset. For instance, if certain demographic groups or transaction types are disproportionately excluded during the anonymization process, the resulting dataset may not accurately reflect the true distribution of fraudulent activities across the population. Future research should strive to obtain datasets that are representative of the broader financial landscape while adhering to strict privacy and ethical standards.

Moreover, the use of machine learning models for real-time fraud detection raises significant ethical concerns surrounding consumer privacy and data protection. While the goal of such systems is to prevent financial losses and protect consumers from fraudulent activities, the continuous monitoring and analysis of transaction data may be perceived as an invasion of privacy. Financial institutions must be transparent about their data collection and usage practices, providing clear opt-in and opt-out mechanisms for consumers. Additionally, robust security measures must be implemented to prevent unauthorized access to sensitive consumer data and to ensure that the fraud detection system itself is not vulnerable to exploitation by malicious actors.

Balancing the need for effective fraud prevention with the imperative to respect consumer privacy is a delicate task that requires ongoing collaboration between financial institutions, regulators, and consumer advocacy groups. As the use of machine learning in fraud detection becomes more widespread, it is crucial to establish clear guidelines and best practices that prioritize both security and privacy. This study, while limited in its scope, aims to contribute to this important conversation by demonstrating the potential of advanced machine learning techniques in combating

financial fraud while also highlighting the ethical considerations that must be addressed as these technologies are deployed in real-world settings.

Thirdly, unsupervised learning techniques had many limitations that reduced their practical application. Lastly, the research relied on historical data, which posed a significant challenge because it made it difficult for the model to adapt and predict new patterns.

1.8 Research Relevance

This research is relevant in the current financial world, where everything has become digital and financial fraud has increased exponentially. The digitization of the sector has made it a lucrative target for malicious actors bent on accomplishing unethical objectives. The traditional methods of detecting fraud cannot address the digital world's challenges. Relying on these methods has proven to be a catastrophic mistake due to their palpable shortcomings (Ryman-Tubb et al., 2018; Ngai et al., 2011). More advanced methods are needed to accomplish this task. Machine learning addresses the modern demands for more efficient and accurate fraud detection systems (Ryman-Tubb et al., 2018).

This research addresses the speed, accuracy, and reliability of fraud detection systems. Speed affects reliability because slow detections lead to missed opportunities, meaning the system becomes unreliable. Accuracy affects reliability because trustworthiness depends on accurate assessments. Using supervised and unsupervised techniques enables the research to identify more complex fraud-associated scenarios. The focus on imbalanced data is essential because it allows ML models to address fraud in cases where data is scarce (Phua et al., 2010; Popat and Chaudhary, 2018).

This research is crucial to the financial sector because it demonstrates how ML can be integrated with existing systems to enhance fraud detection. The hybrid approach explored in this research makes systems more adaptable to emerging threats (Botchey et al., 2020; Lokanan, 2023). This research has the potential to reshape the current standards guiding financial crime prevention. These standards could likely be revised to ensure institutions implement practices that guarantee safety. Thus, this research is not merely a theoretical contribution to existing literature. It provides practical guidelines for combating financial fraud (Chaquet-Ulldemolins et al., 2022).

Chapter 2: Literature Review

This literature review examines the current state of research on fraud detection using machine learning techniques, with a specific focus on applications in the financial industry. The review covers various aspects of fraud detection, including supervised learning approaches, unsupervised learning approaches, hybrid models, and the application of these techniques in the Kenyan context. By synthesizing the key findings and identifying research gaps, this review aims to provide a strong foundation for the current study.

Supervised learning methods have been widely applied in fraud detection, particularly in the context of credit card transactions. Afriyie et al. (Afriyie et al., 2023) demonstrated the effectiveness of supervised learning methods, such as Support Vector Machine (SVM), in detecting fraudulent patterns with high accuracy (Albashrawi, 2016). Bhattacharyya et al. (Bhattacharyya et al., 2011) compared the performance of several supervised classifiers, including logistic regression, support vector machines, and random forests, on credit card fraud data. They found that random forests achieved the highest accuracy, with a 94% true positive rate and a 3% false positive rate. However, the study also highlighted the challenge of imbalanced class distributions, as fraudulent transactions typically constitute a small minority of the dataset. Dhankhad et al. (Dhankhad et al., 2018) further confirmed the effectiveness of supervised methods, particularly ensemble classifiers, in detecting credit card fraud, achieving accuracy rates above 90%.

Unsupervised learning techniques, such as clustering and anomaly detection, have also been explored for fraud detection. These methods are particularly useful for identifying hidden patterns and novel fraud schemes in unlabeled data. Choi and Lee (Choi and Lee, 2018) examined the application of clustering techniques in fraud detection, highlighting their efficiency in identifying hidden patterns within transactions. Schreyer et al. (Schreyer et al., 2017) applied a combination of clustering and anomaly detection techniques to identify suspicious insurance claims, demonstrating the ability of unsupervised learning to detect novel fraud schemes that may have been missed by supervised classifiers. Similarly, Phua et al. (Phua et al., 2004) employed minority oversampling and anomaly detection methods to tackle the class imbalance problem

in automobile insurance fraud, showcasing the potential of unsupervised techniques to enhance fraud detection performance. Bakumenko and Elragal (Bakumenko and Elragal, 2022) further explored anomaly detection and how ML models have made such detections easier and faster.

Recent research has increasingly focused on developing hybrid approaches that combine the strengths of supervised and unsupervised learning for fraud detection. Zhu et al. (Zhu et al., 2018) proposed a hybrid model that integrates supervised random forests with unsupervised anomaly detection based on autoencoder networks, achieving superior performance compared to standalone supervised or unsupervised methods. Similarly, Zhang et al. (Zhang et al., 2019) developed a hybrid approach that combines supervised gradient boosting with unsupervised clustering for credit card fraud detection, demonstrating improved accuracy and robustness to evolving fraud patterns. The empirical evidence suggests that hybrid approaches can offer significant benefits in fraud detection by leveraging the complementary strengths of supervised and unsupervised learning, achieving higher detection accuracy, lower false positive rates, and greater resilience to evolving fraud tactics.

In Kenya, financial fraud, mainly through mobile transactions, has risen as more people adopt mobile banking services such as M-Pesa. Despite the growing threat, research on fraud detection systems tailored for mobile transactions is limited. This gap is particularly significant given Kenya's position as a global leader in mobile banking with over 83% penetration. While existing literature covers various ML techniques and their applications in different financial contexts, there remains a critical research gap in developing integrated hybrid models that combine supervised learning, unsupervised learning, and ensemble methods specifically optimized for Kenya's mobile money ecosystem. Furthermore, previous studies have not adequately addressed the practical implementation challenges of deploying such models in production-ready web and mobile applications, nor have they sufficiently tackled the severe class imbalance problem inherent in Kenyan mobile money fraud detection, where legitimate transactions vastly outnumber fraudulent ones.

This literature review has examined the current state of research on fraud detection using ML techniques, focusing on supervised learning approaches, unsupervised learning approaches, hybrid models, and their application in the Kenyan context.

The review has highlighted the effectiveness of supervised learning methods, particularly ensemble classifiers, in detecting credit card fraud, as well as the potential of unsupervised techniques, such as clustering and anomaly detection, in identifying novel fraud schemes. Furthermore, the review has emphasized the benefits of hybrid approaches that combine supervised and unsupervised learning for enhanced fraud detection performance.

However, the review has also identified significant research gaps, particularly in the context of fraud detection for mobile transactions in Kenya. There is a need for integrated hybrid models that are specifically optimized for Kenya's mobile money ecosystem and can effectively address the challenges of class imbalance and practical implementation in production-ready web and mobile applications. Addressing these gaps requires a comprehensive approach that not only advances theoretical frameworks but also provides practical, implementable solutions for financial institutions facing substantial losses due to increasingly sophisticated fraud schemes. The current study aims to fill these research gaps by developing and evaluating advanced machine learning algorithms for detecting financial fraud in the Kenyan context, with a primary focus on supervised and ensemble methods while also exploring the potential of unsupervised learning techniques.

Chapter 3: Methodology

3.1 CRISP-DM

This study used the CRISP-DM methodology for fraud detection. It is a model that explains the phases of data mining and how to ensure success in each stage. The model has six phases, which can be modified to align with the addressed task. The key to having an effective model is to have a balanced dataset. The SMOTE technique was used to ensure that class imbalances were effectively addressed and that the model was more efficient. While the random forest classifier was the main algorithm, other models were implemented to compare them with this algorithm. The CRISP-DM methodology simplifies data mining by enhancing knowledge of each process. In this way, it becomes relatively easy to build dependable models. The datasets used in this study had features such as transaction amounts, locations, and customer details. The goal was to create an ML model that accurately identifies fraudulent transactions within a system. The CRISP-DM methodology covers everything from understanding the business to deploying the model.

3.2 Business Understanding

The first phase involves a clear definition of the business problem: Kenyan financial institutions need a system to detect fraudulent transactions in real time, particularly in the context of mobile payments, which have become a target for cybercriminals. Machine learning is increasingly being used in the financial sector to detect fraud. It has been established that its use addresses most of the challenges institutions face regarding fraudulent transactions. The research model can be adapted to address credit/debit card transactions. Fraud prevention requires security measures geared towards modeling the differences between balanced and imbalanced data. From the business perspective, modeling the differences depends on the researcher's ability to identify how algorithmic models can be integrated into security systems to advance security features.

The fraud detection model prioritizes:

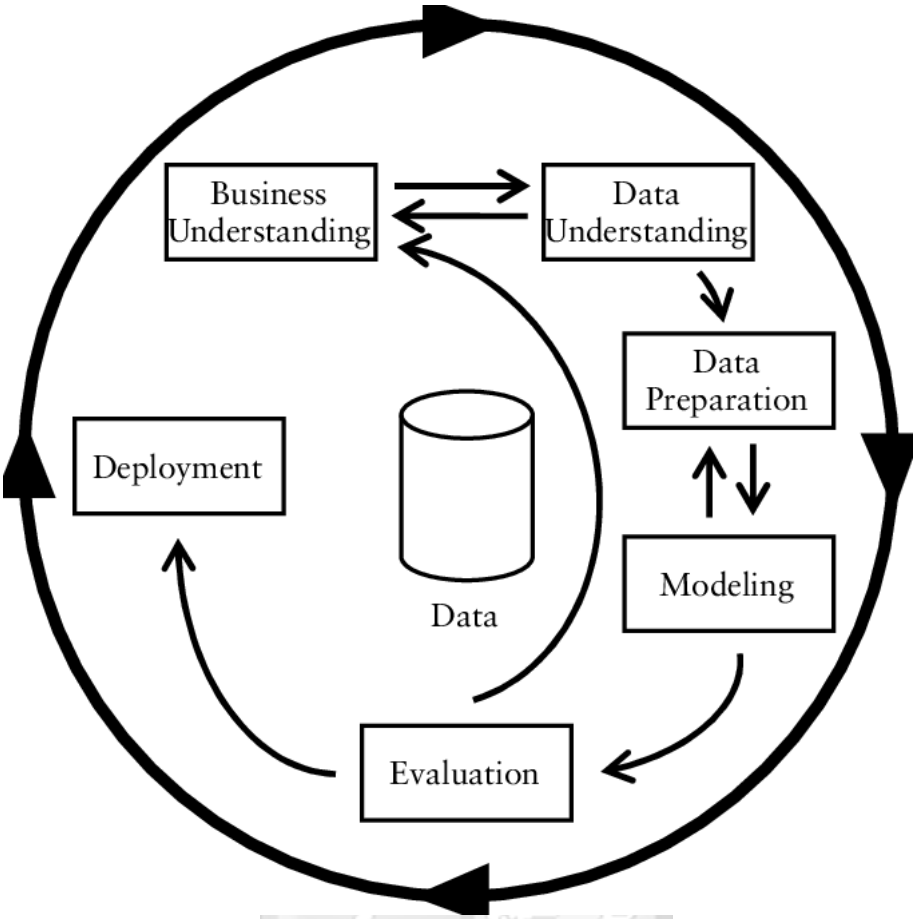


Figure 3.1: CRISP-DM process model

(a) High recall (sensitivity in identifying fraud) while minimizing false positives.

(b) Scalability to handle increasing transaction volumes and different types of financial fraud.

Key performance indicators included the detection rate, false positive rate, and the model's ability to adapt to emerging fraud patterns. The objective was to develop a model that accurately classified transactions as fraudulent or non-fraudulent, minimizing false positives while maximizing fraud detection accuracy.

3.3 Data Understanding

The study assessed how machine learning can prevent financial fraud in the country. The first step in understanding the data involved verifying whether the collection process captured the correct data. This required evaluating the data for accuracy and completeness, ensuring that everything needed to be recorded was recorded and that the models aligned with the available data. Secondly, the research reviewed the captured observations to ensure they conformed to requirements. Thirdly, the study checked the data outliers to idealize the link between the captured data and the strategies for handling the outliers. In addition, the study examined other relationships in the data, such as trends and patterns. Lastly, the understanding of the data led to identifying key responses to the questions raised based on the expected data.

The dataset is expected to include a variety of features, such as:

(a) **Transaction Amount:** The monetary value associated with each transaction. This feature is crucial for identifying unusual or large transactions that may indicate fraud. For instance, if a customer typically makes small transactions but suddenly initiates a large one, it could flag the customer for further investigation.

(b) **Timestamp:** The exact time at which each transaction occurred, represented in a format that allows for temporal analysis. Analyzing the timing of transactions can help identify patterns that may suggest fraud. For example, transactions occurring at unusual hours or in rapid succession could be flagged.

(c) **Transaction Frequency:** The number of transactions performed within a specified time frame, which can help identify patterns of activity. Monitoring how often transactions occur can help detect anomalies. An increase in transaction frequency from a specific account could indicate fraudulent activity.

(d) **Ratio of Incoming to Outgoing Transactions:** A measure comparing the number of transactions received versus those sent, useful for assessing financial activity balance. This ratio can help identify accounts with unusual activity patterns. For example, an account with an unusually high ratio of incoming to outgoing transactions might suggest money laundering.

(e) **Number of Failed Transactions:** The count of transactions that were not successfully completed, which can indicate potential issues or fraud attempts. A high number of failed transactions from an account could indicate attempts at fraudulent activity, such as trying to bypass security measures.

(f) **Frequency of the Same Amount of Transactions:** The occurrence rate of transactions with identical amounts, which might suggest repetitive or automated transactions. Repeated transactions of the same amount could suggest automated or fraudulent activities, such as money laundering schemes.

(g) **Customer Details:** Information about the individuals involved in transactions, including first name, last name, and job title, which can provide context about transaction legitimacy. While not directly indicative of fraud, these details can provide context when combined with other features. For instance, a job title that doesn't align with the transaction behavior could raise suspicions.

(h) **Geographic Information:** Detailed location data, including longitude, latitude, city, state, and zip code, useful for spatial analysis and identifying regional trends. Geographic information can help identify transactions that occur in unexpected locations, which might indicate fraud. For example, if a transaction originates from a location far from the customer's usual activity area, it could be flagged.

(i) **Transaction Details:** Specifics about each transaction, such as transaction date and Unix time, which facilitate temporal analysis and sequencing of events. These details facilitate temporal analysis and can help identify patterns or anomalies in transaction timing.

(j) **Transaction Category:** The type or classification of each transaction (for instance, purchase, transfer), which helps in understanding transaction context. Knowing the type of transaction (for instance, purchase, transfer) can help assess its legitimacy based on the customer's typical behavior.

(k) **Customer Account Age:** This feature can be particularly useful in fraud detection. New accounts are often more susceptible to fraud, as fraudsters may create them for malicious purposes. Monitoring transactions from newly created

accounts can help identify potential fraud early on. Conversely, older accounts with a history of legitimate transactions are generally less likely to be involved in fraud, unless there is a sudden change in behavior.

(l) **Is Fraud:** A binary variable indicating whether a transaction is fraudulent, serving as the target variable for the study.

In this study, the researcher examined the impact of using ML algorithms to prevent fraud in the financial sector. A critical step in understanding data involved checking to ensure that the process of collecting data captured the right data. The target variable, "is fraud," was imbalanced, posing a high chance for misclassifications without adequate strategies to prevent the same from happening. The rarity of such transactions meant that the algorithm encountered more legitimate transactions than fraudulent ones.

3.4 Data Preparation

Data preparation was among the most commonly used methods to address data imbalance. It addressed this imbalance by adding more fraudulent transactions to the data and reducing the number of legitimate transactions. This balancing ensured that the model remained accurate and that learning was efficient. The model became more efficient and precise after this imbalance was addressed. A crucial step in this process involved measuring how well the model handled class imbalances. The two metrics used were the Balanced Classification Rate (BCR) and the Matthews Correlation Coefficient.

The SMOTE technique was used to address imbalances but had many downsides that the researcher needed to consider. Its main downside was that it increased the odds of class overlap. This reality meant the model struggled to process similar data in the future because its learning was limited to the previous data. Studies on SMOTE indicate that it can be effective at addressing data imbalance. One study showed that accuracy improved by 2-4% compared to other approaches for classifying data.

Many new methods have emerged and shown lots of promise in improving classification. The issue with these new methods, like Adaptive Synthetic Sampling (ADASYN) and Ranking Minority Over-sampling Boost (RAMO), is that they add more problems

to the already complex equation. These methods slow algorithms due to the number of learning cycles they need to classify data. Researchers have observed that combining multiple classifiers is better than relying on one when dealing with data imbalance. Cost-sensitive learning was applied to adjust the model.

Data preparation involved cleaning, transforming, and preparing the dataset for modeling. This phase handled missing values, converted categorical data into a usable format, and created new features. Feature engineering was used to generate new features to improve the model's ability to detect fraud. The preprocessing of the captured data incorporated a multi-step process. First, the researcher imported the necessary libraries, packages, resources, and datasets. The step ensured that the collected data was retrieved from the source to allow the analysis process to proceed while verifying the availability of all required variables. The researcher identified and established the independent variables in the second step. The next step involved extracting the dependent variable based on the identified independent variables. The goal of identifying the independent and the dependent variables was to streamline the data analysis process.

3.5 Exploratory Data Analytics

Exploratory Data Analysis (EDA) was conducted to understand the distribution of features, detect outliers, and identify relationships between variables. The research found key insights from EDA, including the transaction amount followed a right-skewed distribution, with a few high-value transactions; certain geographic locations exhibited higher instances of fraud; fraudulent transactions tended to occur more frequently during specific times of the day. The EDA process demonstrated that the failure to collect sufficient data hindered the proper completion of subsequent steps in the research process. After collecting the required data, the next step involved identifying the feature variables and understanding them. In this step, the researcher established relevant variables likely to affect outcomes the most. Next, the EDA process included dataset cleaning procedures. Cleaning involved checking for null values, renaming column names, transforming data types, and removing duplicates. This step was critical since it focused on solving data preprocessing issues that could have hindered analysis outcomes. Next, categorical features such as city, state, zip code, job, and

category were encoded using one-hot encoding to convert them into numerical format suitable for machine learning algorithms. Furthermore, new features were created to capture additional information. For instance, temporal features like the day of the week, month, and hour were extracted from the transaction date. Geographic features such as the distance from a central point were computed to capture location-based patterns. Numerical features were standardized using StandardScaler to ensure they had a mean of zero and a standard deviation of one, which helped improve the performance of many machine learning algorithms. In identifying correlated variables, the process included taking the data through a correlation matrix to idealize the vital relationships between variables and the probable impact of this correlation on analysis outcomes. Additional steps included choosing the most befitting statistical methods and visualizations that assisted in analyzing the data.

3.6 Machine Learning Modeling

The modeling phase involved selecting, training, and tuning machine learning algorithms to build the fraud detection model. The process of modeling machine learning to coincide with the need to use related algorithms to prevent fraudulent financial activities in Kenya followed several predefined steps. In the first step, the research identified and collected the required data. The step also involved understanding the requirements of the machine learning model. Data preprocessing and preparation followed while testing and training tools were used to enhance accuracy and performance. Selecting the right learning model hinged on the nature of the problem at hand and its complexity. In this study, which pursued modeling a machine learning language to identify how related algorithms could solve financial fraud in Kenya, the right model depended on experimenting with multiple approaches to determine the best fit. The research employed various machine learning models, with a particular focus on:

(a) *Supervised learning:* Algorithms such as Random Forests, Gradient Boosting Machines (GBMs), and Neural Networks were used to detect fraudulent transactions.

(b) *Unsupervised learning:* Clustering techniques such as K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) were used to

detect anomalous patterns in the transaction data.

(c) *Ensemble methods:* The study explored ensemble methods that combined multiple models to improve fraud detection accuracy.

Real-time fraud detection was achieved by integrating these models with stream processing platforms capable of handling real-time data inflows. Model performance was evaluated based on accuracy, precision, recall, and F1-score.

3.6.1 Random Forest

At the onset, the study began with the Random Forest Classifier because of its ability to handle high-dimensional data, robustness to overfitting, and capability to capture complex interactions between features. A Random Forest model consisted of an ensemble of decision trees, and its output was based on the aggregated predictions of multiple individual trees. The model is formalized as follows:

Given a set of training data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, where x_i represents the input features and y_i is the target (fraudulent or non-fraudulent), Random Forest works by:

(a) *Bootstrapping:* Creating multiple datasets $\mathcal{D}_b$ from the original dataset $\mathcal{D}$ by random sampling with replacement.

(b) *Tree construction:* For each tree T_b, a subset of features $\mathcal{F}b \subseteq \mathcal{F}$ is randomly selected, and the tree is trained to minimize a loss function, typically Gini impurity or entropy, defined as:
Gini impurity: $G(t) = 1 - \sum i = 1^k p_i^2$

where p_i is the proportion of instances belonging to class i at a node t.

(c) *Prediction:* For a new transaction x, each tree T_b outputs a prediction $y_b(x)$, and the final prediction is the majority vote for classification or the average for regression:
$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b(x)$

where B is the total number of trees.

After preprocessing—cleaning, feature engineering, handling imbalance—the ML model took as input vectors transaction attributes such as location, amount, frequency, and so on. The output was expected to be a binary label representing whether the transaction was classified as fraud or not.

3.6.2 Gradient Boosting Machine

Gradient Boosting trains weak learners sequentially, with each learner correcting the residual errors of its predecessor. The objective is to minimize a differentiable loss function by iteratively adding weak learners. Essentially, this is a sequential model that constructs weak learners (decision trees) to enhance prediction accuracy. The model is represented as follows:

Let the training dataset be $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$. The prediction at the $m - th$ stage for a transaction x is:

$$\hat{y}^{(m)}(x) = \hat{y}^{(m-1)}(x) + \alpha \cdot h_m(x)$$

where:

(a) $\hat{y}^{(m)}(x)$ is the prediction at stage m,

(b) $\hat{y}^{(m-1)}(x)$ is the prediction from the previous stage,

(c) $h_m(x)$ is the $m-th$ weak learner trained on the residuals of the previous learners, and

(d) α is the learning rate.
The loss function $L(y, \hat{y})$ is minimized over all data points, and the weak learner is chosen to reduce the residuals.
$h_m = \arg\min_h \sum_{i=1}^{n} L(y_i, \hat{y}^{(m-1)}(x_i) + \alpha \cdot h(x_i))$
For classification tasks, common loss functions include:

(e) Binary cross-entropy
$L(y, \hat{y}) = -\left[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\right]$

The input is the same as Random Forest mentioned earlier. The model outputs a probability score representing the likelihood of fraud. Based on a threshold (typically 0.5), the transaction is classified as fraudulent or non-fraudulent.

3.6.3 K-Means Clustering

K-Means is an unsupervised learning algorithm that attempts to partition the data into k clusters, where each point belongs to the cluster with the nearest centroid. It is useful for identifying anomalous transactions by grouping them into normal and anomalous clusters. K-Means clustering is formalized with the following equation: Given a set of n data points $(x_1, x_2, \ldots, x_n)$, where each data point is an m-dimensional feature vector, the objective of K-Means is to minimize the sum of squared distances between data points and their assigned cluster centers:

$$J = \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

where:

(a) S_i represents the set of points assigned to cluster i,

(b) μ_i is the centroid of cluster i,

(c) $\|x_j - \mu_i\|$ is the Euclidean distance between point x_j and centroid μ_i.

The prepared transaction data is transformed into a feature space that highlights patterns. In the output, cluster labels are observed, where transactions in clusters that deviate significantly from normal patterns (anomalous clusters) are flagged as suspicious and potentially fraudulent.

3.6.4 DBSCAN

DBSCAN is another unsupervised learning algorithm that groups data points based on density. It is effective for fraud detection as it identifies outliers (anomalous transactions). The density-based clustering method is suitable for finding clusters of irregular shape and detecting noise (outliers, which could indicate fraud). The

following formulation applies to DBSCAN models:

For a set of points ($X = x_1, x_2, \ldots, x_n$), DBSCAN groups points into clusters based on two parameters: ϵ (the neighborhood radius) and $MinPts$ (the minimum number of points required to form a dense region).

(a) *Core point:* A point is a core point if there are at least $MinPts$ points within a radius ϵ.

(b) *Directly density-reachable:* A point p is directly density-reachable from a core point q if p lies within ϵ of q.

(c) *Density-reachable:* A point is density-reachable if it can be reached by a chain of directly density-reachable points.

The input consists of preprocessed transaction data with normalized features. In the output, transactions are categorized as belonging to clusters (normal transactions) or as outliers (potentially fraudulent transactions).

3.6.5 SMOTE for Handling Imbalanced Data

SMOTE generates synthetic samples for the minority class (fraudulent transactions) to balance the dataset. It is formalized as follows:

For each minority class instance x_i, SMOTE generates synthetic examples by randomly interpolating between x_i and one of its nearest neighbors x_{nn}:

$$x_{\text{new}} = x_i + \lambda(x_{nn} - x_i)$$

where $\lambda \in [0, 1]$ is a random number.

3.6.6 Other Models

In addition to the primary models discussed earlier (Random Forest, Gradient Boosting Machine, K-Means Clustering, and DBSCAN), the study also considered other widely used machine learning models for fraud detection. These models included Logistic Regression, SVM, and ANN.

Logistic Regression is a popular choice for binary classification tasks, such as distinguishing between fraudulent and non-fraudulent transactions. It estimates the probability of an instance belonging to a particular class based on a linear combination of input features. Logistic Regression is known for its simplicity, interpretability, and ability to handle high-dimensional data.

SVM are another powerful class of supervised learning algorithms. Support Vector Machines (SVMs) aim to find the optimal hyperplane that maximally separates the different classes in the feature space. They can handle both linear and nonlinear decision boundaries through the use of kernel functions. SVMs are effective in handling high-dimensional data and have shown good performance in various fraud detection tasks.

ANN are inspired by the structure and function of biological neural networks. Artificial Neural Networks (ANNs) consist of interconnected nodes (neurons) organized in layers, which learn to map input features to output predictions through a process called backpropagation. Deep learning architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have gained popularity in fraud detection due to their ability to capture complex patterns and temporal dependencies in data.

After selecting the models to be evaluated, the researcher employed the commonly used 80-20 split mechanism to divide the dataset into training and testing sets. The 80-20 split involves randomly partitioning the dataset, where 80% of the data is used for training the models and the remaining 20% is used for testing their performance on unseen data. This split allows for a reliable assessment of the models' generalization ability and helps prevent overfitting.

3.7 Machine Learning Model Evaluation and Optimization

In this research, several performance metrics have been employed to evaluate the effectiveness of classification models. Each metric offers a distinct perspective on model performance, and their combined use provides a comprehensive assessment. The metrics discussed include Accuracy, AUC, Precision, and Recall

3.7.1 Accuracy

Accuracy is the proportion of correct predictions to the total number of predictions. For instance, if the built model makes 100 predictions and 90 of them are correct, the accuracy is 90%. This metric provides an overall sense of how often the model is right. Although accuracy gives an overall measure of model performance, it may be misleading in cases of imbalanced datasets, where a model might predominantly predict the majority class while neglecting the minority class.

The accuracy was computed using Equation 1.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP represents *True Positives*, TN represents *True Negatives*, FP represents *False Positives*, and FN represents *False Negatives*.

3.7.2 Area Under the Curve

The AUC, quantifies the model's ability to discriminate between positive and negative classes across various threshold levels. The Receiver Operating Characteristic (ROC) curve plots the true positive rate against the false positive rate. A higher AUC value, nearing 1.0, indicates a strong discriminative capacity, whereas an AUC around 0.5 suggests performance close to random guessing.

3.7.3 Recall

Recall, also referred to as sensitivity, is the proportion of correctly identified positive cases to the total actual positives. For instance, if the model successfully detects 80 out of 100 actual positive fraud cases, the recall is 80%. High recall is particularly important in scenarios where missing a positive case can have significant consequences, ensuring that the model captures most of the relevant instances.
Recall is computed using Equation 2:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

where TP represents *True Positives* and FN represents *False Negatives*

3.7.4 Precision

Precision, calculated using Equation 3, measures the ratio of true positive predictions to the total number of positive predictions made by the model. For example, if a model identifies 50 instances as positive and 40 of these are true positives, the precision is 80%. This metric is critical when the cost of false positives is high, as it reflects the reliability of the model's positive classifications.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

where TP represents *True Positives* and FP represents *False Positives*.

3.7.5 F1-Score

Calculated as the harmonic mean of precision and recall, the F1-score provides an integrated measure of model performance by balancing false positives and false negatives. This metric is particularly useful when dealing with uneven class distributions since it captures both Type I and Type II errors. By combining precision and recall into one statistic, it allows for a straightforward comparison of models that trade off these errors differently. The F1-score is computed using the formula Equation 4:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Collectively, these metrics provide a multifaceted view of model performance. Accuracy offers a general measure of correctness; AUC assesses the model's overall discriminative power; precision evaluates the trustworthiness of positive predictions; and recall indicates the model's ability to capture all positive instances. The integration of these metrics was vital for a robust evaluation of the applied classification models, which involved a complex and imbalanced dataset.

3.7.6 Hyperparameter Tuning with GridSearchCV

Hyperparameter tuning is a crucial step in optimizing the performance of machine learning models. It involves systematically searching for the best combination of

hyperparameters that maximizes the model's predictive accuracy and generalization ability. In this study, we employed the GridSearchCV class from the scikit-learn library (Pedregosa et al., 2011) to perform an exhaustive search over a specified parameter grid, evaluating each combination of hyperparameters using cross-validation. For the Random Forest model, we tuned the following hyperparameters:

(a) `n_estimators`: The number of trees in the forest. We considered values of 100, 200, and 300 to balance model complexity and computational efficiency.

(b) `max_features`: The number of features to consider when looking for the best split. We explored using the square root (`sqrt`) and the logarithm (`log2`) of the total number of features.

(c) `max_depth`: The maximum depth of each tree in the forest. We considered depths of 10, 20, and 30 to control the complexity of the individual trees.

(d) `min_samples_split`: The minimum number of samples required to split an internal node. We explored values of 2, 5, and 10 to prevent overfitting and ensure a minimum number of samples in each leaf.

(e) `min_samples_leaf`: The minimum number of samples required to be at a leaf node. We considered values of 1, 2, and 4 to control the minimum size of the leaf nodes and prevent overfitting.

For the XGBoost model, we tuned the following hyperparameters:

(a) `n_estimators`: The number of boosting rounds or trees to build. We explored values of 50, 100, and 200 to find the optimal model complexity.

(b) `max_depth`: The maximum depth of each tree. We considered depths of 3, 5, and 7 to control the complexity of the individual trees and prevent overfitting.

(c) `learning_rate`: The step size shrinkage used in updating the weights. We explored values of 0.01, 0.1, and 0.2 to control the contribution of each tree and find the optimal learning rate.

(d) `subsample`: The fraction of samples to be used for fitting each tree. We considered values of 0.6, 0.8, and 1.0 to introduce randomness and prevent overfitting.

(e) `colsample_bytree`: The fraction of columns (features) to be used for fitting each tree. We explored values of 0.6, 0.8, and 1.0 to introduce randomness and reduce overfitting.

(f) `reg_alpha`: L1 regularization term on weights. We considered values of 0, 0.1, and 1 to control the sparsity of the model and prevent overfitting.

(g) `reg_lambda`: L2 regularization term on weights. We explored values of 0, 0.1, and 1 to control the regularization strength and improve generalization.

For the Logistic Regression model, we tuned the following hyperparameters:

(a) `C`: The inverse of the regularization strength. We explored values of 0.1, 1, and 10 to find the optimal balance between fitting the training data and avoiding overfitting.

(b) `solver`: The algorithm used in the optimization problem. We considered the `liblinear` and `saga` solvers, which are suitable for small and large datasets, respectively.

To evaluate the performance of each hyperparameter combination, we employed a 5-fold cross-validation strategy. This approach involves splitting the training data into 5 equal-sized subsets, using 4 subsets for training and the remaining subset for validation. The process is repeated 5 times, with each subset serving as the validation set once. Cross-validation helps to assess the model's performance on unseen data and reduces the risk of overfitting.

The objective of the hyperparameter tuning was to maximize the F1-score metric, which is the harmonic mean of precision and recall. The F1-score is particularly suitable for imbalanced fraud detection datasets, as it balances the model's ability to correctly identify fraudulent transactions (recall) while minimizing false positives (precision). By optimizing the F1-score, we aimed to develop models that can effectively detect fraud while maintaining a low false positive rate.

After evaluating all hyperparameter combinations using GridSearchCV, we selected the best combination for each model based on the highest average F1-score across the cross-validation folds. These optimal hyperparameters were then used to train the final models on the entire training set and evaluate their performance on a separate test set. This approach ensures that the models are tuned to generalize well to unseen data and provides a reliable estimate of their real-world performance. By conducting a thorough hyperparameter tuning process using GridSearchCV and cross-validation, we aimed to develop robust and accurate fraud detection models that can effectively identify fraudulent transactions in the Kenyan digital banking context. The selected hyperparameters strike a balance between model complexity, generalization ability, and computational efficiency, ensuring the practicality and reliability of the developed models.

3.8 Deployment

The production environment enabled the model to be used for its intended purpose. It processed live transactions and flagged potential fraud. The pre-trained model was deployed on a mobile and/or web application, focusing on making predictions and alerting users when fraud was detected. The application was deployed to a closed group of users—banking sector practitioners. The precision and accuracy of predictions depended on the model's training and preparation outcomes. After successful training and preparation, the deployment process integrated the model into a production environment (mobile and/or web) and subjected it to real digital financial transactions to validate real-time performance. The model scored each transaction in real time and flagged suspicious transactions for further investigation. Managing these outcomes ensured that, once the machine learning model was subjected to real-time data, it produced coherent results aligned with its intent: detecting, preventing, and alerting fraudulent mobile transactions in the digital banking sector. Furthermore, continuous performance monitoring was required to verify the model's efficacy. Regular updates and retraining were undertaken to adapt to changing fraud patterns and maintain high detection rates.

Chapter 4: Discussion of Results

Kenyan mobile money transaction data formed the basis for assessing the efficacy of ML algorithms. Preprocessing steps included data cleaning, feature extraction, and scaling. The training and evaluation focused on precision, accuracy, recall, and F1 score.

In addition to these evaluation metrics, hyperparameter optimization played a critical role in enhancing model performance. As discussed in Section 3.7.6, the GridSearchCV technique was employed to systematically search over a predefined set of hyperparameters for each model, using cross-validation to assess their performance. The detailed explanation of the hyperparameter tuning process, including the specific hyperparameters tuned for Random Forest, XGBoost, and Logistic Regression, can be found in Section 3.7.6. Initial results indicated that Random Forest and Gradient Boosting performed better compared to Logistic Regression, achieving a high accuracy rate.

4.1 Machine Learning Modeling

4.1.1 Model Performance Before Hyper Parameter Tuning

The performance of various machine learning models was evaluated on the dataset to assess their ability to detect fraudulent transactions. Before applying hyperparameter tuning, Random Forest, Logistic Regression, and Gradient Boosting were implemented, and their performance metrics were analyzed. While the study primarily focused on

Metric	Random Forest	Logistic Regression	Gradient Boosting
AUC	0.995 ± 0.004	0.781 ± 0.004	0.942 ± 0.010
Accuracy	93%	73%	90%
Precision (Class 1)	0.32	0.12	0.26
Recall (Class 1)	0.29	0.65	0.52
F1-Score (Class 1)	0.30	0.20	0.34
Log-Loss	0.297	0.532	0.334
ROC-AUC	0.758	0.780	0.820
Precision-Recall AUC	0.255	0.323	0.297

Table 4.1: Performance Comparison of Different Models

supervised and ensemble methods, unsupervised learning techniques such as K-Means and DBSCAN were also explored for their potential to identify patterns of fraudulent behavior. However, due to the exploratory nature of this analysis and the challenges of evaluating clustering performance on highly imbalanced datasets, the results of the clustering experiments are not included in the main discussion of the study's findings.

The Random Forest model exhibited a high AUC score of 0.995 ± 0.004, indicating perfect separation between the classes during the training phase. However, its classification report revealed significant limitations in detecting fraudulent transactions (class 1). While the precision and recall for legitimate transactions (class 0) were exceptionally high (0.96 and 0.97, respectively), the model completely failed to identify fraudulent activities, with both precision and recall at 0.32 and 0.29, respectively. This resulted in a macro average F1-score of 0.63, highlighting the model's moderate balance in performance across classes. The Log Loss value of 0.297, ROC-AUC score of 0.758, and Precision-Recall AUC of 0.255 further illustrated the challenge in detecting minority class instances.

Logistic Regression similarly achieved an AUC score of 0.781 ± 0.004 during training. The classification report demonstrated superior performance compared to Random Forest in identifying fraudulent transactions, albeit with notable shortcomings. The precision for fraudulent transactions reached 0.12, but the recall remained moderate at 0.65, yielding an F1-score of 0.20. Conversely, the model effectively identified legitimate transactions, with a precision of 0.98 and recall of 0.73. The overall accuracy was 73%, while the macroaverage F1-score was 0.52. Logistic Regression produced a log loss of 0.532 and an ROC-AUC score of 0.780. However, the Precision-Recall AUC of 0.323 reflected limited capability in distinguishing fraudulent transactions within an imbalanced dataset.

The Gradient Boosting model yielded a training AUC score of 0.942 ± 0.010, signifying slightly reduced performance compared to Random Forest and Logistic Regression. Its classification report revealed better balance across classes, with a recall of 0.52 and precision of 0.26 for fraudulent transactions, resulting in an F1-score of 0.34. For legitimate transactions, the recall was 0.92 and precision was 0.97, respectively. The overall accuracy of 90% and macroaverage F1-score of 0.65 suggested moderate efficacy in detecting fraudulent activities. The Log-Loss value of 0.334, ROC-AUC

score of 0.820, and Precision-Recall AUC of 0.297 demonstrated the model's robustness relative to its peers but still highlighted room for improvement.

4.1.2 Performance Improvements After Hyperparameter Tuning

Hyperparameter tuning was performed to optimize the models and improve their ability to detect fraudulent transactions effectively. This process involved fine-tuning key parameters to enhance predictive accuracy and reduce false negatives.

Model	Class	Precision	Recall	F1-Score	Accuracy
Logistic Regression	0	0.98	0.73	0.84	
	1	0.12	0.65	0.20	**0.73**
Random Forest	0	0.96	0.97	0.97	
	1	0.33	0.29	0.31	**0.93**
XGBoost	0	0.85	0.80	0.82	
	1	0.81	0.86	0.83	**0.83**
Adjusted Threshold	0	0.93	0.66	0.77	
	1	0.74	0.95	0.83	**0.81**

Table 4.2: Model Performance Metrics After Tuning

The optimized XGBoost model emerged as a suitable contender with improved balance in performance measures. The classification report showed a precision of 0.81 and a recall of 0.86 for fraudulent transactions, resulting in an F1-score of 0.83. For genuine transactions, the recorded precision and recall values were 0.85 and 0.80, respectively, contributing to an overall accuracy of 83%. The macro average F1-score improved to 0.83, reflecting greater balance in predictions across classes. Additionally, the Log-Loss value of 0.185 indicated increased stability in the model.

The optimized Logistic Regression model demonstrated notable improvement in identifying fraudulent transactions. The precision for the minority was at 0.12, although the recall remained at 0.65, resulting in an F1-score of 0.20. Legitimate transactions maintained high precision and recall at 0.98 and 0.73, respectively. The overall accuracy was 73%, with a macro average F1-score of 0.52. The model's Log-loss increased slightly to 0.533, reflecting a trade-off between improved fraud detection and minor reductions in certainty for legitimate transactions.

The Random Forest model, after tuning, showed minimal improvement in detecting fraudulent transactions. Precision and recall for the minority class remained at 0.33 and 0.29, respectively, highlighting persistent challenges in handling data imbalance. However, the performance for legitimate transactions remained strong, with a precision of 0.96 and a recall of 0.97. The overall accuracy was 93%, with a macro average F1-score of 0.64. The log loss value was reduced slightly to 0.265, indicating slight gains in prediction confidence without substantial advancements in fraud detection.

4.1.3 Effect of Adjusted Decision Threshold

In addition to hyperparameter tuning, the decision threshold for classification was adjusted to enhance fraud detection capabilities. Table 3 below shows the performance metrics with the adjusted threshold. The adjusted model became more efficient at

Metric	**Adjusted Model**
Accuracy	81%
Precision (Class 1)	0.74
Recall (Class 1)	0.95
F1-Score (Class 1)	0.83
Macro Avg. F1-Score	0.80

Table 4.3: Performance Metrics with Adjusted Thresholds

identifying fraudulent transactions. It had a recall of 0.95, meaning it could identify 95% of actual cases accurately, but it still made minor mistakes (precision of 0.74). While still low, the F1 score of 0.83 for the minority class shows the model is improving and getting better than before. Legitimate transactions had high precision (0.93) and recall (0.66) scores. These high numbers were not surprising because these events make up the bulk of the transactions in the financial sector. The challenge is in detecting the rare events. The overall accuracy dropped to 81%, suggesting that while it is performing, areas for improvement exist to increase the accuracy. The macro average F1-score was 0.65, a decent score in many respects considering the challenge of classifying the transactions. The results highlight the necessary trade-off in such situations: to detect more fraud, the overall accuracy has to be compromised.

4.1.4 Visual Analysis of Model Performance

To further illustrate the performance of the machine learning models, visual representations of the evaluation metrics are provided below. Figure 4.2 presents the ROC curve and Precision-Recall curve for the best-performing model, demonstrating its ability to discriminate between fraudulent and non-fraudulent transactions. Figure 4.1 shows the class distribution before and after applying the SMOTE technique, highlighting the impact of addressing class imbalance on the model's performance. These

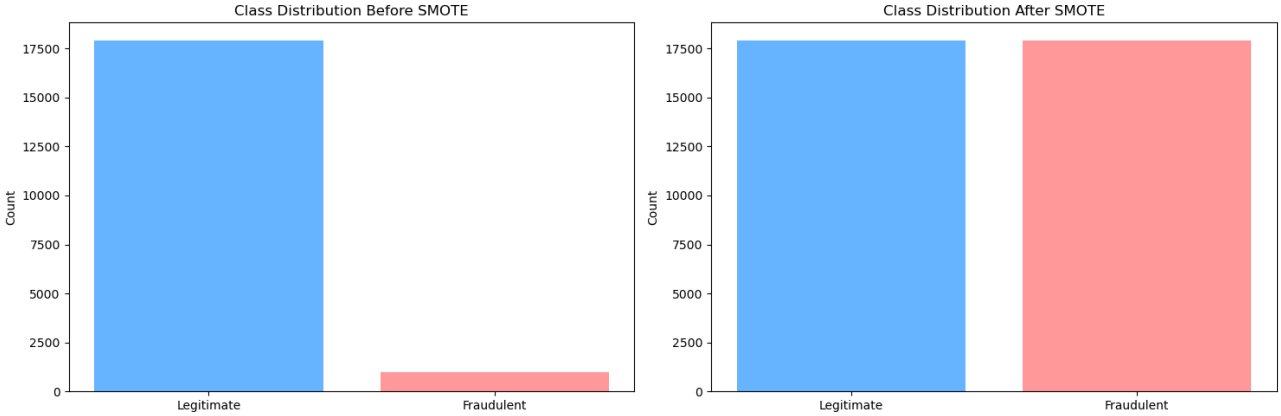


Figure 4.1: Class Distribution before and after SMOTE

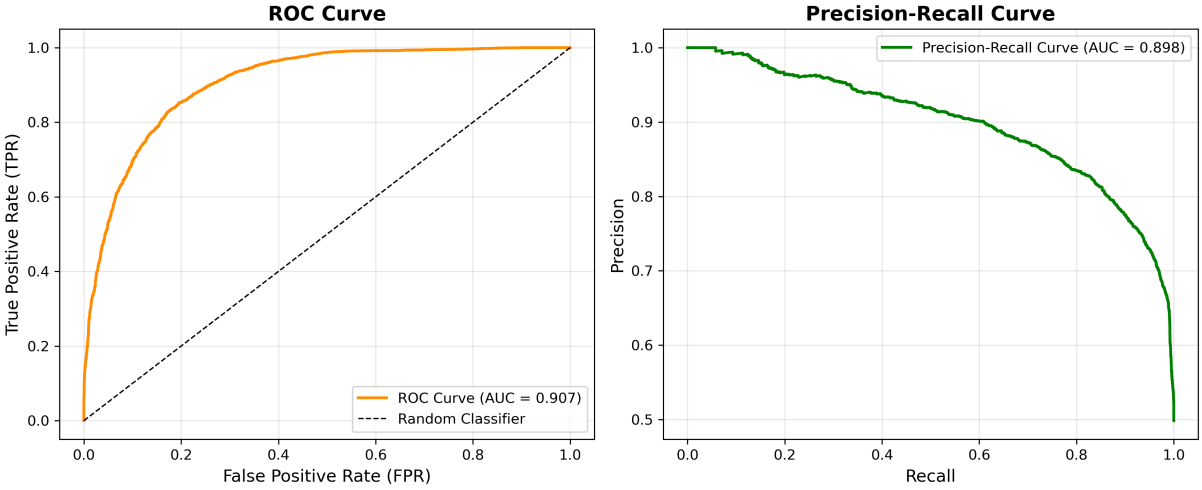


Figure 4.2: ROC Curve and Precision-Recall Curve for the best-performing model

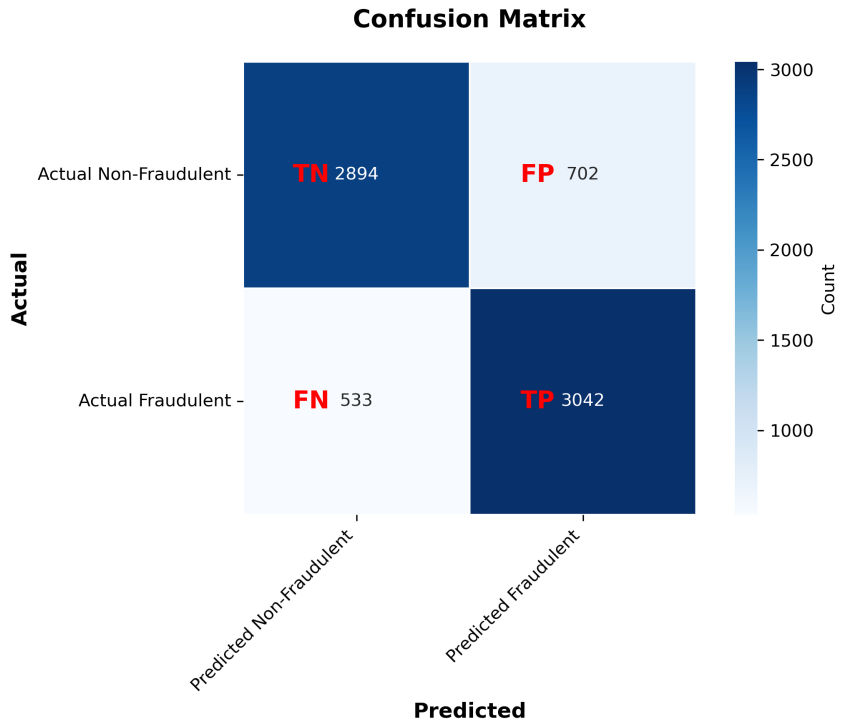


Figure 4.3: Confusion Matrix for the best-performing model

visual representations provide a comprehensive overview of the model's performance and complement the quantitative results discussed in the previous subsections. The ROC curve and Precision-Recall curve (Figure 4.2) demonstrate the model's ability to balance sensitivity and specificity, while the class distribution plots (Figure 4.1) highlight the effectiveness of the SMOTE technique in addressing class imbalance. Finally, the confusion matrix (Figure 4.3) provides a detailed breakdown of the model's predictions and misclassifications, enabling a deeper understanding of its performance.

4.1.5 Summary

The research explored the different models used in detecting financial fraud. The results increased understanding of their strengths and areas where they are less likely to generate accurate results. Before tuning, Gradient Boosting showed the most balanced performance. Random forest and logistic regression struggled with minority classes. The two are generally accurate but cannot detect rare fraudulent cases. After

tuning, XGBoost was the most efficient, especially since adjustments were made to its decision threshold. These tunings suggest that models become more effective when strategically tuned, and their thresholds are adjusted to detect fraud. The results confirmed the research was on the right path regarding building a secure and efficient fraud detection system.

4.2 Summary

The study revealed challenges and opportunities with using ML models to detect financial fraud in Kenyan mobile transactions. The results show that traditional models are inefficient at detecting actual fraud. The issue with financial fraud is that it is less common than genuine ones, resulting in data imbalance. Accuracy in fraud detection requires addressing this problem, which the two models (random forests and logistic regression) struggle with. Gradient Boosting and XGBoost yielded better results and became more efficient when adjustments were made to their parameters and thresholds.

The random forest model struggled with detecting minority classes. Other studies on the model have made similar observations, meaning its efficiency with imbalanced datasets depends on the preprocessing techniques employed (Ashtiani and Raahemi, 2021). Logistic regression was moderately successful in detecting fraud but missed many actual fraudulent transactions. This limitation is related to the fact that it (the model) excels with labeled data but struggles with imbalanced datasets. In sum, the two models exhibited some notable successes in detecting fraud but had some limitations.

Hyperparameter tuning resulted in notable changes in the performance of Gradient Boosting and XGBoost. The results demonstrated that XGBoost was the most efficient and accurate fraud detection tool. Its ability to effectively address uneven data and adjust to varied changes makes it ideal for detecting fraud. This finding is similar to what other researchers have observed about XGBoost (Bakumenko and Elragal, 2022). The results suggest that having an advanced model is not enough. The desired results are achieved through modifying parameters to ensure they work efficiently in a given environment.

The findings suggest a high probability of ensuring mobile money is safe in Kenya.

Based on the results, XGBoost is best suited to address the complexities of financial transactions. Imbalanced datasets often make it challenging to detect and prevent financial fraud accurately. With models like XGBoost, the odds of incorrect flags are reduced, meaning only a few instances can occur. Accuracy and efficiency are not just about correctly identifying unusual patterns within the system but also achieving much with minimal resource consumption.

Integrating these models into financial institutions increases the chances of developing systems that can detect fraud in real time. Besides precision in fraud detection, such systems are fast and efficient, using computing resources economically. For example, incorporating XGBoost into a fraud detection system would increase its chances of detecting all frauds. The study shows that these models can be made better with additional features. Only relevant features should be added to avoid introducing new complexity issues.

This study contributes to understanding the challenges within the Kenyan financial sector. Many studies on this subject mostly focus on the European context, which differs from how people use mobile money in Kenya. The research observed that some models were effective in other contexts and the country. XGBoost generated results similar to those other researchers observed in their studies (Afriyie et al., 2023). The universal effectiveness of these models does not mean they just get implemented as they are. The study emphasizes the need to localize these models as fraud differs across regions.

Room for improvement exists with the proposed models. Future research should focus on how unsupervised learning techniques can advance these proposed models. Specifically, researchers can explore how using autoencoders can enhance efficiency and accuracy. There is also the need to incorporate real-time capabilities to capture data that can be used to improve them. Tailoring these models to align with the challenges in the industry increases understanding of the financial sector. Another essential part is understanding why the system flags some transactions as fraudulent and others as genuine. Explainable AI helps with the endeavor, enhancing understanding of the system's operation.

The study demonstrates that ML models can efficiently prevent fraud in the financial sector. The models have to be modified and adjusted regarding how they make

decisions to enhance their efficiency and accuracy at spotting fraud. These findings highlight the need for robust and efficient systems to detect real-time fraud. These systems should be built to address the country's financial sector challenges. Future research should use these findings to develop more efficient systems.

Chapter 5: Conclusion, Recommendations and Future Work

5.1 Conclusion

Institutions must address the digitization challenges with everything going digital in the financial sector. Financial fraud cases have increased, with many expecting the trend to remain the same due to the high value of financial assets. The study addressed this issue by exploring how machine learning can enhance security. Various ML models were tested and evaluated to establish their capacity to combat financial fraud. Of the models, XGBoost emerged as the most efficient when adjusted to how it made decisions. Many fraudulent transactions were detected accurately without compromising privacy. The results indicate that ML models can prevent financial fraud in the mobile money sector. The models will enable institutions to be more proactive than reactive, which has mostly been the trend. Proactiveness in this context means being one step ahead in securing transactions from malicious actors. ML models can improve and become more efficient with time, ensuring systems are always protected against new threats. With ML models, financial institutions are assured of consumer trust in their services and the growth of their revenues. Cybercrimes are increasing and evolving in the financial sector, emphasizing the need for researchers to continue developing ML models. While effective in some aspects, the current models lack the accuracy and speed needed in the industry. Future studies should concentrate on these two areas and how to improve real-time fraud detection. Past studies suggest that the solution to the current problem could be combining supervised and unsupervised learning methods, making this another essential area for research. In conclusion, ML models appear to be the most effective solution to the fraud challenges in the Kenyan financial sector.

5.2 Recommendations

Based on the findings and discussion in this study, several recommendations can be made to enhance the application of machine learning for fraud detection in Kenyan digital banking. First, financial institutions should consider integrating a hybrid approach that combines supervised and unsupervised techniques. Such an approach can balance the strengths of each method—supervised models offer high precision with

labeled data, while unsupervised models can uncover hidden patterns in unlabeled datasets. This integration should be supported by regular updates to address emerging fraud patterns and evolving cyber threats. Second, institutions are encouraged to invest in advanced data preprocessing and feature engineering techniques. In particular, refining the application of SMOTE and exploring alternative sampling methods could further mitigate the challenges posed by imbalanced datasets. Additionally, establishing continuous monitoring mechanisms that use real-time data streams will improve the responsiveness of the fraud detection system. It is also recommended that organizations prioritize cross-departmental collaboration, involving both technical teams and fraud management specialists, to ensure that the insights generated by ML models are effectively translated into actionable policies. Finally, regulatory bodies should provide frameworks and incentives for banks to adopt these innovative techniques, thus promoting a more secure financial ecosystem. These measures, collectively, are expected to not only reduce financial losses but also to reinforce consumer trust and strengthen the overall security infrastructure within the Kenyan financial sector.

5.2.1 Piloting the Application with Banks and Fintechs

To ensure the successful implementation and adoption of the proposed fraud detection application, it is crucial to conduct a pilot study in collaboration with banks and fintech companies. The pilot study will provide valuable insights into the real-world performance of the application, user acceptance, and potential areas for improvement. The following steps outline a comprehensive plan for piloting the application, taking into account the need for stakeholder engagement, iterative refinement, and post-deployment evaluation.

Stakeholder Engagement The first step in piloting the application is to identify and engage with relevant stakeholders in the banking and fintech sectors. This includes:

(a) Identifying potential partner banks and fintech companies that have a strong interest in fraud detection and are willing to participate in the pilot study. Key criteria for selection may include the size of the institution, the volume of digital transactions processed, and their current fraud detection capabilities.

(b) Conducting initial meetings with key decision-makers, such as Chief Information Security Officers (CISOs), Chief Technology Officers (CTOs), and Heads of Fraud Prevention, to present the proposed application and its benefits. These meetings should focus on demonstrating how the application can enhance existing fraud detection processes, reduce financial losses, and improve customer trust.

(c) Establishing a stakeholder advisory board comprising representatives from participating organizations to provide guidance and feedback throughout the pilot study. The advisory board should include members with diverse expertise, such as fraud analysts, data scientists, and customer service representatives, to ensure a comprehensive perspective on the application's performance and usability.

Pilot Study Design Once the stakeholders are engaged, the next step is to design the pilot study. This involves:

(a) Defining clear objectives and success metrics for the pilot study, such as fraud detection accuracy, false positive rates, and user satisfaction. These metrics should be aligned with the key performance indicators identified in the business understanding phase (Section 3.2) and should be measurable and relevant to the participating organizations.

(b) Determining the scope of the pilot study, including the duration, the number of participating organizations, and the types of transactions to be monitored. The duration should be sufficient to capture a representative sample of transactions and allow for meaningful evaluation of the application's performance. The types of transactions monitored should cover a range of scenarios, including mobile money transfers, card payments, and online banking transactions.

(c) Developing a detailed implementation plan, including timelines, resource allocation, and responsibilities. The plan should outline the steps for integrating the fraud detection application with the participating organizations' existing systems, as well as the process for data collection, monitoring, and reporting. It should also specify the roles and responsibilities of the research team, the advisory board, and the participating organizations.

(d) Establishing data sharing and security protocols to ensure the confidentiality and integrity of sensitive financial information. This should involve implementing secure data transfer mechanisms, anonymizing customer data, and adhering to relevant data protection regulations such as the Kenyan Data Protection Act (2019).

Training and Support To facilitate the smooth adoption of the application, it is essential to provide comprehensive training and support to the pilot study participants. This includes:

(a) Conducting training sessions for end-users, such as fraud analysts and customer service representatives, to familiarize them with the application's features and functionality. The training should cover topics such as interpreting the application's outputs, investigating flagged transactions, and providing feedback on the system's performance. The sessions should be hands-on and include practical exercises to ensure a thorough understanding of the application.

(b) Providing technical training for IT staff to ensure proper integration with existing systems and troubleshoot any issues that may arise. The training should cover the technical architecture of the fraud detection application, the data integration process, and the procedures for monitoring and maintaining the system. It should also include guidance on how to handle system updates and resolve common technical issues.

(c) Establishing a dedicated support team to address user queries and provide ongoing assistance throughout the pilot study. The support team should consist of both technical experts and domain specialists who can provide guidance on fraud detection best practices and assist with the interpretation of the application's outputs. The team should be accessible through multiple channels, such as email, phone, and a ticketing system, to ensure prompt resolution of issues.

Feedback and Iteration Throughout the pilot study, it is crucial to gather feedback from stakeholders and end-users to continuously improve the application. This involves:

(a) Conducting regular feedback sessions with the stakeholder advisory board to discuss the application's performance, user experiences, and potential enhancements. These sessions should be held at predefined intervals, such as bi-weekly or monthly, and should provide a forum for open discussion and collaborative problem-solving. The feedback gathered should be documented and used to inform the iterative refinement of the application.

(b) Administering user surveys and interviews to gather insights into the application's usability, effectiveness, and impact on fraud detection processes. The surveys should be designed to capture both quantitative and qualitative feedback, covering aspects such as ease of use, perceived accuracy, and the impact on workload. Interviews with key users, such as fraud analysts and customer service representatives, can provide deeper insights into the application's strengths and weaknesses.

(c) Analyzing the collected feedback and prioritizing improvements based on their potential impact and feasibility. The feedback should be systematically reviewed and categorized based on common themes and the level of urgency. The research team, in collaboration with the advisory board, should prioritize the improvements based on their potential to enhance the application's performance, user satisfaction, and alignment with the overall objectives of the pilot study.

(d) Iteratively refining the application based on the feedback received and releasing updates to the pilot study participants. The prioritized improvements should be implemented in a phased manner, with each update accompanied by detailed release notes and user guidance. The updated versions of the application should be thoroughly tested before deployment to ensure stability and performance. The iterative refinement process should continue throughout the pilot study to ensure continuous improvement and adaptation to evolving fraud patterns.

Evaluation and Scaling Upon completion of the pilot study, it is important to evaluate the results and determine the next steps for scaling the application. This includes:

(a) Assessing the application's performance against the predefined success metrics and comparing it to the baseline fraud detection processes. The evaluation should involve a comprehensive analysis of the application's accuracy, precision, recall, and F1-score, as well as its impact on false positive rates and investigation times. The results should be compared to the performance of the existing fraud detection processes to quantify the improvement achieved through the application.

(b) Conducting a cost-benefit analysis to determine the potential Return on Investment (ROI) of implementing the application on a larger scale. The analysis should consider factors such as the reduction in fraud losses, the efficiency gains in fraud investigation processes, and the potential improvements in customer satisfaction and trust. The costs associated with the application's development, deployment, and maintenance should be weighed against the projected benefits to determine the overall ROI.

(c) Developing a roadmap for scaling the application, including timelines, resource requirements, and potential partnerships. Based on the results of the pilot study and the cost-benefit analysis, a detailed plan for scaling the application should be developed. The roadmap should outline the steps for expanding the application's deployment to additional banks and fintech companies, as well as the resources required for ongoing development, support, and maintenance. Potential partnerships with technology providers, industry associations, and regulatory bodies should be explored to facilitate wider adoption.

(d) Sharing the pilot study results and lessons learned with the wider banking and fintech community to promote the adoption of advanced fraud detection techniques. The research team should actively engage with industry stakeholders through conferences, workshops, and publications to disseminate the findings of the pilot study and showcase the benefits of the fraud detection application. By sharing the lessons learned and best practices, the team can contribute to the advancement of fraud detection capabilities in the Kenyan financial sector and foster a culture of collaboration and innovation.

By following this comprehensive plan for piloting the application with banks and fintechs, the research team can gather valuable insights, refine the application based on real-world feedback, and lay the foundation for wider adoption in the industry. The collaborative approach, involving active stakeholder engagement, iterative refinement, and rigorous evaluation, will help ensure that the fraud detection application meets the needs of its intended users and contributes to the overall goal of reducing financial fraud in the Kenyan context.

5.3 Future Work

Future research should build upon the current study by further refining and expanding the machine learning framework for fraud detection. One promising direction is the exploration of deep learning techniques, such as autoencoders and recurrent neural networks, which may capture complex temporal patterns and nonlinearities more effectively than traditional models. Future work should also investigate the integration of explainable AI methods to better understand and justify the decision-making process of the detection system. This transparency is essential for regulatory compliance and for gaining the confidence of stakeholders. In addition, researchers are encouraged to explore real-time model deployment using edge computing and cloud-based platforms, which could significantly reduce detection latency in live transaction environments.

While the exploratory analysis of clustering methods, such as DBSCAN and K-Means, suggests their potential for uncovering patterns of fraudulent behavior, further research is needed to fully investigate their applicability and integrate them with existing fraud detection systems. Future work could focus on developing more robust evaluation metrics for clustering performance on imbalanced datasets and exploring hybrid approaches that combine clustering with supervised and ensemble methods.

Expanding the dataset to include a wider range of fraudulent activities—beyond those commonly observed in credit and mobile transactions—would also enhance model robustness and generalizability. Furthermore, collaborative studies involving multiple financial institutions can facilitate the creation of a shared, anonymized dataset that captures diverse fraud scenarios, leading to improved model performance. Finally, longitudinal studies assessing the long-term impact of these technological interventions on fraud reduction and consumer behavior will provide deeper insights

and guide continuous improvements in digital banking security.

Bibliography

Abdallah, A., Maarof, M. A., and Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, 68:90–113.

Achary, R. and Shelke, C. J. (2023). Fraud detection in banking transactions using machine learning. In *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, pages 221–226. IEEE.

Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., and Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*, 6:100163.

Al-Hashedi, K. G. and Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, 40:100402.

Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. *Journal of Data Science*, 14(3):553–569.

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., and Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. *Applied Sciences*, 12(19):9637.

Ashtiani, M. N. and Raahemi, B. (2021). Intelligent fraud detection in financial statements using machine learning and data mining: a systematic literature review. *Ieee Access*, 10:72504–72525.

Bakumenko, A. and Elragal, A. (2022). Detecting anomalies in financial data using machine learning algorithms. *Systems*, 10(5):130.

Bhattacharyya, S., Jha, S., Tharakunnel, K., and Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision support systems*, 50(3):602–613.

Botchey, F. E., Qin, Z., and 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(8):383.

Carneiro, E. M., Dias, L. A. V., Da Cunha, A. M., and Mialaret, L. F. S. (2015). Cluster analysis and artificial neural networks: A case study in credit card fraud detection. In *2015 12th international conference on information technology-new generations*, pages 122–126. IEEE.

Chaquet-Ulldemolins, J., Gimeno-Blanes, F.-J., Moral-Rubio, S., Muñoz-Romero, S., and Rojo-Álvarez, J.-L. (2022). On the black-box challenge for fraud detection using machine learning (ii): nonlinear analysis through interpretable autoencoders. *Applied Sciences*, 12(8):3856.

Choi, D. and Lee, K. (2018). An artificial intelligence approach to financial fraud detection under iot environment: A survey and implementation. *Security and Communication Networks*, 2018(1):5483472.

Da'u, A. and Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. *Artificial Intelligence Review*, 53(4):2709–2748.

Delamaire, L., Abdou, H., and Pointon, J. (2009). Credit card fraud and detection techniques: a review. *Banks and Bank systems*, 4(2).

Dhankhad, S., Mohammed, E., and Far, B. (2018). Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, pages 122–125. IEEE.

Gyamfi, N. K. and Abdulai, J.-D. (2018). Bank fraud detection using support vector machine. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 37–41. IEEE.

Hilal, W., Gadsden, S. A., and Yawney, J. (2022). Financial fraud: a review of

anomaly detection techniques and recent advances. *Expert systems With applications*, 193:116429.

Iyer, D., Mohanpurkar, A., Janardhan, S., Rathod, D., and Sardeshmukh, A. (2011). Credit card fraud detection using hidden markov model. In *2011 World Congress on Information and Communication Technologies*, pages 1062–1066. IEEE.

Lokanan, M., Tran, V., and Vuong, N. H. (2019). Detecting anomalies in financial statements using machine learning algorithm: The case of vietnamese listed firms. *Asian Journal of Accounting Research*, 4(2):181–201.

Lokanan, M. E. (2023). Predicting mobile money transaction fraud using machine learning algorithms. *Applied AI Letters*, 4(2):e85.

Muiruri, K. (2023). Absa bank loses sh107 million to fraudsters. `https://www.businessdailyafrica.com/bd/corporate/companies/absa-bank-loses-sh107-million-to-fraudsters-4414226`. Accessed: (2024, June 20).

Munyua, J. M. (2013). *Operational response strategies to payment card fraud by commercial banks in Kenya*. PhD thesis, University of Nairobi.

Nation, T. (2013). Rogue staff have stolen sh1.5 billion from banks. `https://nation.africa/kenya/news/rogue-staff-have-stolen-sh1-5-billion-from-banks--863926`. Accessed: (2024, June 20).

Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., and Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, 50(3):559–569.

Patil, S., Nemade, V., and Soni, P. K. (2018). Predictive modelling for credit card fraud detection using data analytics. *Procedia computer science*, 132:385–395.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Phua, C., Alahakoon, D., and Lee, V. (2004). Minority report in fraud detection: classification of skewed data. *Acm sigkdd explorations newsletter*, 6(1):50–59.

Phua, C., Lee, V., Smith, K., and Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.

Popat, R. R. and Chaudhary, J. (2018). A survey on credit card fraud detection using machine learning. In *2018 2nd international conference on trends in electronics and informatics (ICOEI)*, pages 1120–1125. IEEE.

Raj, S. B. E. and Portia, A. A. (2011). Analysis on credit card fraud detection methods. In *2011 International Conference on Computer, Communication and Electrical Technology (ICCCET)*, pages 152–156. IEEE.

Ryman-Tubb, N. F., Krause, P., and Garn, W. (2018). How artificial intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence*, 76:130–157.

Schreyer, M., Sattarov, T., Borth, D., Dengel, A., and Reimer, B. (2017). Detection of anomalies in large scale accounting data using deep autoencoder networks. *arXiv preprint arXiv:1709.05254*.

West, J. and Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security*, 57:47–66.

Zeng, Y. and Tang, J. (2021). Rlc-gnn: An improved deep architecture for spatial-based graph neural network with application to fraud detection. *Applied Sciences*, 11(12):5656.

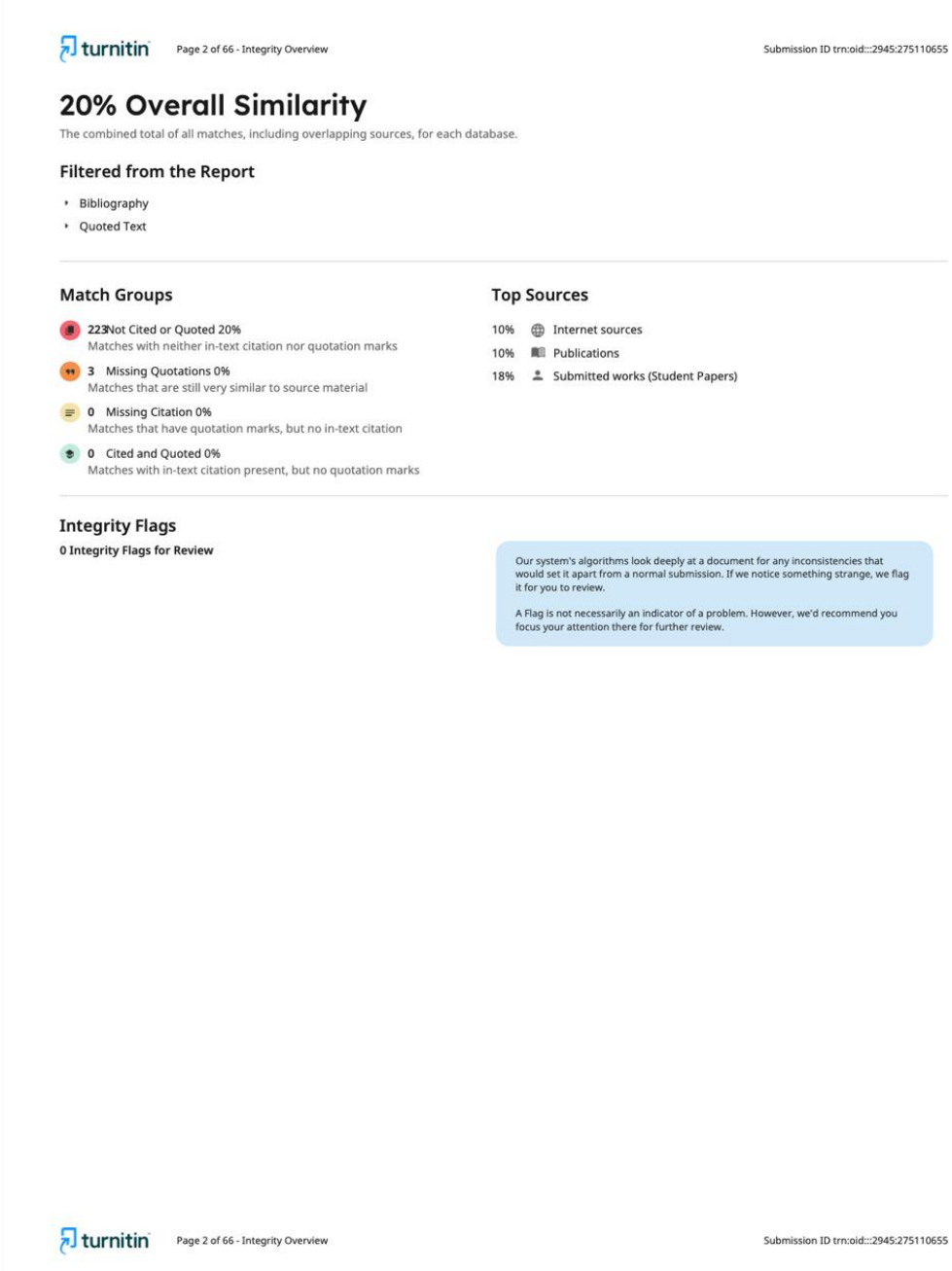
Zhang, D. and Zhou, L. (2004). Discovering golden nuggets: data mining in financial application. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34(4):513–522.

Zhang, X., Han, Y., Xu, W., and Wang, Q. (2019). Hoba: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences*, 557:302–316.

Zhu, C., Zhang, H., Qin, Y., and Xu, L. (2018). A hybrid model for credit card fraud detection based on neural network and random forest. In *2018 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA)*, pages 185–190. IEEE.

Appendices

Appendix A: Similarity Report



turnitin Page 2 of 66 - Integrity Overview Submission ID trn:oid:::2945:275110655

20% Overall Similarity

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

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

223 Not Cited or Quoted 20%
Matches with neither in-text citation nor quotation marks

3 Missing Quotations 0%
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

10% Internet sources
10% Publications
18% 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.

turnitin Page 2 of 66 - Integrity Overview Submission ID trn:oid:::2945:275110655

Appendix B: Ethical Clearance Release Letter

Strathmore UNIVERSITY

25th March 2025

Mr Imende Michael,
michael.edward@strathmore.edu

Dear Mr Imende,

RE: Applying Machine Learning to Enhance Fraud Detection in Kenyan Digital Banking

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2740/25.** The approval period is from **25th March 2025 to 24th March 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu