ADDIS ABABA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES
SCHOOL OF INFORMATION SCIENCE

MOBILE BANKING AND MOBILE MONEY FRAUD DETECTION USING MACHINE LEARNING ON BANKS IN ETHIOPIA

BY

DANIEL MANAYE

ID# GSE/2902/12

February 2024

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES
SCHOOL OF INFORMATION SCIENCE

MOBILE BANKING AND MOBILE MONEY FRAUD DETECTION USING MACHINE LEARNING ON BANKS IN ETHIOPIA

A thesis submitted to the school of graduate studies of Addis Ababa University in partial fulfilment of the requirements for the degree of Master of Science in information science and systems (information systems specialization)

BY

DANIEL MANAYE

ID# GSE/2902/12

February 2024

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

MOBILE BANKING AND MOBILE MONEY FRAUD DETECTION USING MACHINE LEARNING ON BANKS IN ETHIOPIA

BY

DANIEL MANAYE

Name and signature of Members of the Examining Board

Workshet Lamenew (Ph.D.)	_____	_____
Advisor	**Signature**	**Date**
_____	_____	_____
Examiner	**Signature**	**Date**
_____	_____	_____
Examiner	**Signature**	**Date**

February 2024

Addis Ababa, Ethiopia

Declaration

This thesis not previously been submitted for any degree and is not being concurrently submitted in candidature for any degree in any university. I declare that this thesis entitled "MOBILE BANKING AND MOBILE MONEY BANKING FRAUD DETECTION USING MACHINE LEARNING ON BANKS IN ETHIOPIA" is a result of my own investigation, except where otherwise stated. I have undertaken the study independently with the guidance and support of my research advisor. Other sources acknowledged by citations giving explicit references and a list of references appended.

Signature: _____

Daniel Manaye Kassahun (ID# GSE/2902/12)

This thesis submitted for examination with my approval as university advisor.

Advisor's Signature: _____

Workshet Lamenew (Ph.D.)

Dedication

In loving memory of my dear brother, whose unwavering support for my education from a young age has left an enduring legacy. I am deeply grateful to my beloved family, particularly my mom, whose unwavering encouragement has been a guiding light throughout my educational journey. To my father, whose commitment to education and discipline has shaped my character and aspirations. And to my cherished wife, her solid support has been the foundation of our family's strength and success.

Acknowledgements

My deepest gratitude goes first to the almighty God for helping me in all the ways of my life. In my journey through this study, my advisor, Workshet Lamenew (Ph.D.), has helped me all the way to finalize this thesis. He is, in my view, a democrat and an amazing philosopher who questions all ideas brought to the table, from which I have learned a lot. Thank you once again. My deepest gratitude also goes to my family and to all those who supported me in getting the resources that were required to complete my research without hesitation. I am grateful to all the school of information science instructors who have inspired me to this end and to the staff who have supported me.

Abstract

Fraud is a criminal act with significant societal impacts, particularly in the financial sector. As digitalization expands rapidly in Ethiopia, the financial industry has suffered substantial losses, with an estimated 1.8 billion birr lost to fraud over the past four years. This study aims to explore and evaluate the effectiveness of SVM machine learning technique for detecting fraud in mobile banking and mobile money services within the Ethiopian banking sector. This research employs a quantitative experimental approach to investigate fraud detection in mobile banking and mobile money services using machine-learning models, particularly Support Vector Machines (SVM).

A comprehensive literature review reveals that while mobile banking and mobile money have become essential in East African countries, including Ethiopia, there is a significant gap in research addressing fraud detection in this context. As the digital financial landscape evolves, the threat of fraud is becoming increasingly severe, posing a substantial challenge to the region's financial stability.

This study aims to bridge this research gap by exploring and evaluating the effectiveness of machine learning (ML) techniques for detecting fraud in mobile banking and mobile money services. Utilizing the CRISP-DM framework for data mining, the study apply SVM supervised ML techniques to transaction data from these platforms. To address the class imbalance inherent in the data, under sampling techniques were employed, with the dataset split into training (80%) and testing (20%) sets after the necessary data cleaned and preparation based on the framework selected has been carried out. In this study, the data taken for analysis is the transaction data of mobile banking and mobile money this is because the fraudulent activities on one of the channel may come to the other as the services are having many similarities in nature.

The performance of a Support Vector Machine (SVM) model was assessed using metrics such as Precision, Recall, and Confusion Matrix. Initial findings indicate that the model struggles with the class imbalance, which affects its overall effectiveness but still identify 51% of the fraudulent transactions. Despite these challenges, this study provides valuable insights into the application of machine learning for fraud detection in mobile banking and mobile money services within East African region where such practices are still emerging.

This research contributes to the limited body of knowledge on fraud detection in the rapidly expanding digital financial services sector in East Africa, offering a foundation for future studies and practical applications in the financial section in the region and beyond.

Keywords: Fraud detection, Mobile banking fraud detection, Mobile money fraud detection

Contents

Table of figures

Abbreviations

AI- Artificial Intelligence

ML- Machine learning

SLR-systematic literature review

CFE - Certified Fraud Examiner

ACFE- Association of Certified Fraud Examiners

SVM: Support Vector Machines,

HMM: Hidden Markov Models,

ANN: Artificial Neural Networks,

KNN: K-Nearest Neighbour's

CRISP-DM (CRoss-Industry Standard Process for Data Mining)

DST -Data Science Trajectories

USSD -Unstructured Supplementary Service Data)

SIM-Subscriber Identity Module

NBE-National Bank of Ethiopia

MFIs- Micro Finance Institutions

RDBMS-Database management system

ETB-Ethiopian Birr

PIN- personal identification number

UAT- User acceptance test

VM- virtual machine

PDA -Personal Digital Assistants

PWC- Price water house coopers.

E-payment –Electronic payment

EDA-Exploratory Data Analysis

CHAPTER ONE

1. INTRODUCTION

1.1. Background: Fraud in Mobile Banking and Mobile Money

The banking sector is experiencing rapid growth globally, fueled by globalization and technological advancements. However, this growth also presents new challenges, with financial crimes like fraud posing a significant threat. Fraud erodes public trust in financial institutions and disrupts economic stability. According to the Association of Certified Fraud Examiners (ACFE, 2022), global losses due to fraud exceed $3.6 trillion annually, with the banking sector experiencing the highest number of fraudulent cases and a median loss of $117,000 per case.

In Ethiopia, the banking sector has also experienced remarkable growth. According to the National Bank of Ethiopia (NBE) Quarterly Bulletin 2022/23, there are currently 31 banks in Ethiopia, with 29 being private and 2 state-owned. These banks have opened 314 new branches, bringing the total number of bank branches to 11,281. The total capital rose from Birr 199 billion to Birr 246.7 billion, with private banks holding 63.8% and state-owned banks holding 36.2% of the total.

The role of Micro Finance Institutions (MFIs) in Ethiopia is also significant. NBE Quarterly Bulletin reports that there are 47 MFIs, which mobilized approximately Birr 27.9 billion in saving deposits, showing a 28.8 percent annual growth. The outstanding credit of these institutions increased by 23.7 percent to reach Birr 39.2 billion. The growing presence of MFIs plays a crucial role in promoting financial access, reducing poverty, and fostering wealth creation among low-income groups and micro and small-scale enterprises. The total assets of MFIs also grew by 18.0 percent, reaching Birr 54.4 billion.

Financial crimes in banks involve illegal practices to gain unlawful financial advantages. These crimes include fraudulent deposits, withdrawals, transfers, loan applications, investments, and deceptive offers. Perpetrators can range from individuals and organizations to government-sponsored actors.

Banking institutions are vulnerable to fraud due to the presence of abundant cash resources, attracting various fraudulent activities and embezzlement. The Association of Certified Fraud Examiners (2011) defines fraud as any illegal act characterized by deceit, concealment, or violation of trust, not dependent on threats of violence or physical force. Fraud in the banking sector can be categorized into three main sub-categories: technology-related frauds, KYC-related frauds, and advance-related frauds. It can also be classified into insider fraud (perpetrated by cunning staff), outsider fraud (committed by external parties), and collaborative fraud involving collusion between staff members and external fraudsters (Deepa and Lalita, 2023).

Financial crimes, such as fraud, pose a significant challenge to the banking industry globally. The African continent has substantially faced financial criminal acts, more than other regions. According to PWC's Global Economic Crime Survey in 2016, Africa's reported 'economic' crimes rose 7% over the previous two years, reaching 57%. The worldwide average is 36%. Banking fraud cases have grown significantly concerning. These incidents severely harm banks, customers, stakeholders, and the overall economy financially and non-financially.

The Need for Improved Detection Methods

Given the limitations of traditional or rule based methods and the growing sophistication of fraud, there is a critical need for improved detection methods in Ethiopia's mobile money and mobile banking landscape. Machine learning offers promising potential in this regard. Machine learning algorithms can analyse vast amounts of transaction data to identify complex patterns and anomalies that might be indicative of fraudulent activity. This allows for a more automated and proactive approach to fraud detection, compared to traditional methods.

Interconnected Nature of Fraud in Mobile Money and Mobile Banking

The study focuses on detecting fraud in both mobile money and mobile banking due to the interconnected nature of transactions in fraudulent cases. In a case fraud that has happened and based on the explanations from the bank fraud related experts, said that they have noticed transaction done on the one of the payment channels will be send to the other in most of the circumstances. Whenever fraudulent transactions happens the choice of fraudsters is to send it to another channel and this is to evade detection by the system.

Investigators have also observed that some channels may involve walk-in customers, making it a significant challenge to trace the identity of the fraudsters. The following points are the summary of the rational as why the two payment services are under the same study:

- **Cross-Channel Fraudulent Activity:** Fraudsters often exploit vulnerabilities across mobile money and mobile banking platforms run an attack.
- **Following the Money Trail (tracing movement of money):** Stolen money goes from one channel to the other. Understanding how fraudsters move stolen funds across mobile money and mobile banking allows developing more comprehensive detection methods.

Thus, the study combines both mobile banking and mobile money transactions, analysing the data separately for fraud detection using a machine learning approach. This research specifically targets fraud detection in mobile money and mobile banking transactions in Ethiopia, where these platforms are becoming more interconnected and utilized by customers seamlessly, allowing fraudsters to exploit weaknesses across both channels.

Why mobile money and mobile banking as growing targets for fraud

Mobile money and mobile banking have seen explosive growth in recent years, especially in developing economies like Ethiopia (Mengash & Girma, 2021). This widespread adoption makes a larger pool of potential targets for fraudsters. On the other hand, compared to traditional banking methods, mobile money and mobile banking services perceived as having weaker security measures by fraudsters, which can be due to factors like SIM swap fraud, weak authentication, and limited awareness of the community about the evolving fraud techniques. In this case, a successful mobile money or mobile banking fraud can have a significant impact on the victim, leading to financial loss and potentially a loss of trust in the financial system in the country at large.

1.2. Machine Learning in Fraud Detection

Machine learning and artificial intelligence have become critical in detecting evolving fraud patterns in mobile banking and mobile money transactions. The increasing complexity of fraud tactics necessitates the adoption of machine learning algorithms and real-time

monitoring systems to improve detection and prevention strategies (Oladimeji Kazeem, 2023; Phua et al., 2010).

Given the complex and ever-changing nature of mobile banking fraud, this research focuses on exploring the effectiveness of specific machine-learning algorithms, such as Support Vector Machines (SVMs), for fraud detection. SVMs are well suited for handling high-dimensional data, a characteristic feature of mobile banking transactions. However, SVMs may not perform optimally with imbalanced datasets, where fraudulent transactions are often a minority. On the other hand, Random Forests, known for their ability to identify complex relationships within data and robustness against outliers, offer potential for detecting anomalies indicative of fraud. As additional research on related topics incorporated in the future, this study will focus on utilizing SVM, with rationale provided later on.

To enhance the accuracy of the machine learning models, this research will employ data pre-processing techniques, such as normalization and feature selection, to prepare mobile banking and mobile money transaction data for analysis. The performance of SVMs will be evaluated using metrics such as precision, recall, and ensuring that the models can accurately identify fraudulent activities. Additionally, the identified fraudulent transactions will be cross verified through manual confirmation processes to ensure their validity.

2. Motivation

The study is motivated by the increasing trend towards digitalization in financial services, especially within banking institutions, while concurrently; fraud has also been increasing in parallel. The significant risk of fraud accompanying the expansion of digital banking services is a key driver for this research. Numerous fraud cases currently under investigation at the Ethiopian courts have caught the attention of the authors, as reported in both the news on the media. This highlights the urgency and relevance of examining fraud in the context of digital financial services, not only in a single study but also in continuous and other related research that is expected to be continued in this context. Furthermore, aside from the hurdles posed by pervasive fraudulent activities targeting customers for various reasons, as outlined in the fraud triangle and fraud demand theories, it is evident that fraud persists due to the presence of rationalization, pressure, capability, and opportunity that are not addressed effectively in any other manner. Managing those factors has a big contribution but is normally

possible for the commercial institution as the mandate could go to the government and other stokeholds in society. Furthermore, limited information is available on the nature and scope of e-banking fraud in Ethiopia due to banks' reluctance to officially disclose such data (Yonas, 2020). This lack of transparency delays the development of effective mitigation strategies across the industry.

This research aims to address this critical gap by examining mobile banking and mobile money fraud within the Ethiopian banking context. By analysing existing fraud cases using machine learning detection methods, this research will contribute to a support of the challenges faced by Ethiopian banks to detect fraud on a timely bases for action. The findings will be valuable for developing context-specific solutions and informing future research in this emerging challenge of the banking industry.

3. Statement of the Problem

The rapid growth of digital banking, particularly mobile banking, has revolutionized financial services in Ethiopia. However, this accessibility has also introduced a growing threat of fraud. While the banking industry is heavily, regulated, financial crimes continue to rise globally. A 2023 report by Waleed Hilal et al. estimates annual global losses due to fraud at hundreds of billions of dollars. The World Economic Forum (2018) placed the cost of fraud and financial crime at $2 trillion annually, with private companies spending $8.2 billion on anti–money laundering (AML) controls in 2017, which is often linked to fraud.

In Ethiopia, the situation is similarly being tough challenge as reported by the Ministry of Justice, it was revealed that over the past five years, Ethiopian banks have lost 1.8 billion birr due to fraudulent activities, with attempted frauds amounting to 370 billion birr since 2017. These figures highlight the urgent need for more robust measures to combat fraud in the banking sector.

While some studies have focused on fraud risks in traditional banking channels, such as card fraud (Tsegaye, 2017), there is a lack of research on mobile banking fraud detection using advanced methods like data analytics and machine learning, which are widely used globally (Ali, A. et al., 2022). Existing methods in Ethiopian banks rely heavily on traditional internal controls and reactive measures, such as post-incident audits (Adane T., 2011; Daniel, 2013), rather than proactive, technology-driven solutions.

In other African countries that have experienced significant growth in mobile banking and mobile money services, similar challenges with fraud have been reported (Karanja, 2017; Wanjohi, 2014). Studies from Kenya highlight the weaknesses of internal controls and the absence of data analytics in fraud detection efforts. These challenges are also present in Ethiopia, where mobile banking fraud has become a significant concern.

The rise of mobile banking fraud cases reported by the Ethiopian Federal Police (The Reporter, June 4, 2022) underscores the need for a more proactive approach to fraud detection, leveraging machine learning techniques like anomaly detection and classification algorithms.

Utilizing these technologies could help Ethiopian banks identify suspicious activities more effectively, allowing them to stay ahead of fraudsters.

This study aims to address the critical challenge of increasing mobile banking and mobile money fraud in the Ethiopian banking environment. By leveraging Machine Learning (ML) techniques such as anomaly detection and classification algorithms, Ethiopian banks can potentially develop more sophisticated fraud detection methods. This research explores the feasibility and potential benefits of using ML in this domain. While focus is not on immediate implementation, the findings can contribute valuable insights for developing more effective fraud detection strategies. The research might identify common fraud patterns, inform policy changes, and ultimately help Ethiopian banks take targeted actions to combat fraud.

In conclusion, the research question to investigate in this particular study is:

1. Considering potential data quality limitations in the Ethiopian banking sector, how effective a particular machine learning models in detecting fraudulent transactions in Ethiopian mobile banking and mobile money services to tackle the very tough issue of fraud?
2. How can data imbalances addressed to optimize the performance of machine learning models for fraud detection in the context of Ethiopian mobile banking and mobile money services?

4. Objective of the Study

4.1. General Objective:

The general objective of this research is to quantitatively assess the effectiveness of Machine Learning (ML) techniques for detecting mobile banking and mobile money fraud in the Ethiopian banking context using metrics like accuracy, precision and the like. This study will explore the feasibility and potential benefits of using ML for fraud detection while considering the limitations of data availability in Ethiopia.

4.2. Specific Objective

This study aims to address the following specified research objectives:

1. Conduct a review of relevant research on machine learning techniques specifically applied to mobile banking and mobile money fraud detection.
2. Prepare data for building a machine-learning model for mobile banking and mobile money fraud detection. This includes collecting relevant data, cleaning it to ensure accuracy and consistency, selecting the most appropriate features for model training, and addressing the challenge of data imbalance commonly encountered in fraud detection datasets.
3. To investigate the feasibility and potential benefits of using specific Machine Learning (ML) algorithms, such as SVM for detecting mobile banking and mobile money fraud in the Ethiopian banking context and the research will evaluate the SVM algorithm's performance in identifying fraudulent transactions compared to traditional methods.
4. Implement a supervised Machine Learning algorithm to detect mobile banking and mobile money fraud in the Ethiopian banking context from the available data.
5. Evaluate the performance of the SVM ML algorithm in identifying fraudulent transactions compared to traditional methods employed by Ethiopian banks.

5. Significance of the study

This study contributes to understanding the fraud risk negatively affecting local banking institutions, particularly regarding fraud detection and prevention. It emphasizes the need for enhanced attention to this issue, advocating for the use of machine learning techniques among others. The research test the effectiveness of a selected machine learning technique suggestions for improving the detection of fraudulent activities, facilitating timely responses based on the established rules and regulations of the banking sector.

The findings will provide valuable insights into a machine learning fraud detection methods passing through various challenges of data quality in local context, allowing banks to enhance their prevention measures, making them more proactive and efficient in meeting fraud detection requirements now and in the future.

Moreover, the results of this research will shed light on new use of machine learning for fraud detection, particularly in prevention through machine learning techniques. The study

evaluates a particular machine learning detection methods for the application it in the mobile banking and mobile money digital banking services .

Furthermore, it may serve as a foundational step for further and more in-depth research into fraud detection and prevention within digital banking services using machine-learning techniques, especially considering the current expansion of these services in Ethiopia, as noted in the NBE report.

6. Scope of the study

This study focuses on applying machine learning for fraud detection and prevention in mobile banking and mobile money transactions in Ethiopia in banking environment only. While "fraud" encompasses a wider range of activities explained in the background of this study, but this research is limited to financial frauds involving transactions through mobile banking and mobile money channels. Other types of fraud, such as insurance fraud, are not within the scope.

The rationale for studying both mobile banking and mobile money lies in their interconnectedness for transaction and growing concern of fraud risk in these his particular banking services. Fraudsters often exploit vulnerabilities across both channels, making it essential to analyse them together.

As the financial ecosystem in local banking, industries are moving to a fully digitalized banking service, most transactions conducted via digital channels like payment cards, mobile money, and mobile banking platforms. As a result, local banks will face the same challenges that other banks in the world have already faced. The use of machine learning for fraud detection known to be one way of digital banking fraud detection but on the other hand, it poses several challenges, as will be discussed in the following sections. The aim of this study is to leverage data analytics and machine learning to identify existing fraud patterns in local banking institutions and provide recommendations on how to respond to such fraudulent activities. The data taken from the commercial bank of Ethiopia.

7. Organization of the study

This study focuses on the use of machine learning to detect and prevent banking fraud in relation to mobile banking and mobile money services within selected commercial banks in Ethiopia.

This thesis has presented the introduction part to explain the background information about encounters of fraud and then proceeds to describe the problem statement, research questions, general and specific objectives, significance of the study, scope and limitations. The second chapter focuses on reviewing relevant literature and providing a detailed examination of the concepts related to the fraud in the machine learning perspective.

Chapter 3 of this thesis outlines the methodology of the study and Chapter 4 provides an analysis of the research findings. This section includes a presentation, description, discussion, and analysis of the data and findings as implemented in the data analytics study. Finally, chapter five presents the main findings, summary, conclusions, and recommendations based on the analysis in chapter four. This study concludes by presenting the way forward, key takeaways, and references to the study. For security reasons, the study may keep the data sources anonymous.

CHAPTER TWO

2. Literature review

2.1. Banking fraud landscape

Banking fraud is becoming a pervasive trouble that has become a major write out for financial institutions all over the world. The advancement of technology has made it easier for fraudsters to engage in fraudulent activities through technology, which is also an enabler to the business on the other way. This reality has necessitated the need for more efficient and effective fraud detection mechanisms. Fraud represents a significant problem for governments and businesses and specialized analysis techniques for discovering fraud using them are required. Some of these methods include knowledge discovery in databases (KDD), data mining, machine learning and statistics. They offer applicable and successful solutions in different areas of electronic fraud crimes. (Chuprina, Roman April 2020)

The primary motive behind fraud is financial gain, and banks, being warehouses of liquid cash, which attracts the attention of fraudsters that pose a significant fraud risk (Sanusi, Rameli, & Isa, 2015). Fraudsters also categorized as internal employees, external individuals such as customers and third parties interacting with banks, or a combination thereof (Sanusi, Rameli, & Isa, 2015). Additionally, the fraud diamond theory, building upon Cressey's fraud triangle theory, encompasses the four elements of opportunity, pressure, rationalization, and capability as drivers for individuals to commit fraud, regardless of whether they are internal or external parties. The fulfilment of these factors can lead to fraud, negatively affecting both customers and institutions.

Common Fraud types and trends

The Association of Certified Fraud Examiners (ACFE) reports fraud costs organizations around 5% of their annual revenue globally (2022). As financial transactions move online, new opportunities arise for fraudsters opened as well. The following are some of the common type of frauds, which are in association with mobile banking and mobile money services from the banks:

- **Account Takeover (ATO):** Fraudsters gain unauthorized access to a victim's mobile banking or mobile money account through various methods like phishing or malware.
- **Payment Fraud:** This involves using stolen credentials or social engineering techniques to initiate unauthorized transactions.

- **Social Engineering:** This is showing sophistication of attackers exploiting any available weakness across technology and people. Fraudsters manipulate victims into disclosing personal information or authorizing fraudulent transactions.

2.2. The Significance of Machine Learning in Combating Financial Fraud

This review examines the application of machine learning (ML) for fraud detection in mobile banking and mobile money environments. While reseach on this specific topic are limited, mobile banking and mobile money services are increasingly relevant, particularly in Africa. Traditional fraud detection methods, which rely on manual reviews and static rules, are becoming inadequate due to their slowness, inefficiency, and vulnerability to sophisticated fraudsters (Alshaseh et al., 2017). In contrast, ML algorithms offer a transformative approach capable of identifying complex patterns and anomalies in transaction data, enabling the detection of evolving fraud schemes (Phua et al., 2010).

ML allows for real-time monitoring of transactions, facilitating quicker responses to potential fraud attempts (Kou et al., 2020). This capability is crucial for the instantaneous nature of mobile banking and mobile money transactions, as it enables early intervention that can prevent fraudulent transactions altogether (Alaiwa et al., 2022; Sarma et al., 2019). By continuously analyzing transaction data, ML models can trigger alerts for suspicious activities, allowing banks to take immediate action, such as blocking transactions or contacting customers for verification.

One significant challenge in fraud detection using ML is data imbalance, where fraudulent transactions often constitute a small minority compared to legitimate ones. This imbalance can lead to biased models that overlook fraudulent activities. However, certain ML

algorithms, such as Support Vector Machines (SVM), can effectively handle imbalanced datasets by finding optimal decision boundaries between classes (Oladimeji Kazeem, 2023).

Moreover, ML can analyse vast amounts of data, including transaction logs, customer profiles, and behavioural patterns, to identify anomalies that traditional methods might miss. By learning from historical data on both fraudulent and legitimate transactions, ML models continuously improve their detection capabilities in real-time. This proactive approach allows for early intervention and potentially prevents fraud before it occurs.

According to the ACFE's Occupational Fraud 2022: A Report to the Nations, organizations utilizing proactive data analytics as an Anti-fraud control experience fraud losses that are 47% lower than those that do not employ such analytics. Traditional methods pose significant financial risks to businesses, potentially damaging their profitability and reputation (ACFE, 2022). The advantages of ML—such as its ability to identify complex patterns, facilitate real-time monitoring, and support continuous improvement—underscore its importance in the fight against financial fraud.

In conclusion, the integration of machine learning into fraud detection is not only essential for enhancing the effectiveness of mobile banking and mobile money services but also crucial for safeguarding financial institutions against evolving fraud tactics. Future research should explore further innovations in ML techniques and their applications in diverse banking contexts.

2.3. Challenges and Considerations for Machine Learning-based Fraud Detection

Compared to traditional methods, ML algorithms can learn and adapt from data, allowing them to detect complex patterns and anomalies that may indicate fraud. However, ML-based fraud detection also faces several challenges:

1. Rapidly developing fraud techniques: Fraudsters constantly develop new methods to bypass detection systems. ML algorithms continuously updated with new data to stay ahead of these evolving threats (Ngai et al., 2011).

2. Large volume of data to search for anomalies: Financial institutions generate vast amounts of data, making it difficult to analyse and identify fraudulent patterns efficiently.
3. False positives: ML models can generate false positives, flagging legitimate transactions as fraudulent. This can lead to customer inconvenience and erode trust in the system. Optimizing algorithms and incorporating domain expertise can help reduce false positives (Thampi et al., 2020).
4. Adaptive fraud techniques: As mentioned earlier, fraudsters can adapt their behaviour to avoid detection. Employing anomaly detection algorithms and unsupervised learning techniques can help identify these adaptive patterns (Liu et al., 2016).

Financial institutions hold vast amounts of data crucial for fraud detection, but sharing across departments or with external parties is often limited due to legal restrictions and internal policies (Assefa et al., 2020). Additionally, analysing this data is challenging because:

- **Data fragmentation:** Information siloed, hindering comprehensive analysis.
- **Data Imbalance:** Fraudulent transactions are a small minority compared to legitimate ones. This imbalance of data on genuine and fraudulent transactions can bias ML models towards the majority class (legitimate transactions) and potentially miss fraudulent activities or generate false positives (Oladimeji Kazeem, 2023). Techniques like oversampling or under sampling can address this (Chawla et al., 2002), but these are not the only ways to overcome the data imbalance challenge.
- **Existence of data availability: In** Ethiopian context, the limited information is available on the nature and scope of e-banking fraud in Ethiopia due to banks' reluctance to disclose officially such data (Yonas, 2020). This is also another challenge for this particular research on the mobile banking and mobile money services. Because of these reasons and because the data holding mechanisms are difficult to proceed with in a single research project, only bank data is considered, which is stated in the scope of this study. Therefore, these limitations can hinder the effectiveness of machine learning for fraud detection at large.

2.4. Machine Learning Techniques for Fraud Detection in Banking

This section explores two main categories of machine learning approaches used for fraud detection in banking: supervised learning and unsupervised learning.

2.4.1. Supervised Learning

For identifying known types of banking fraud, supervised learning is frequently used. This approach involves training a machine-learning model on a set of labelled data, where fraudulent transactions identified and labelled as such. The model gains the ability to identify patterns in the data that discriminate between fraudulent and lawful transactions. By comparing incoming transactions to the model's learnt patterns, the model used to detect fresh cases of fraud once it has been trained (Bhattacharya, Bose, & Neogy, 2018).

- **The random forest algorithm:** A popular supervised learning technique that employs multiple decision trees to predict whether a transaction is fraudulent. Each decision tree segments data based on various factors like transaction size or location, ultimately identifying criteria most correlated with fraud (Bhattacharya, Bose, & Neogy, 2018).
- **The support vector machine (SVM)**: algorithm is another well-liked supervised learning approach used in banking fraud detection. This algorithm finds the decision boundary that distinguishes between genuine and fraudulent transactions by mapping data into a high-dimensional space.

2.4.2. Unsupervised Learning

Unsupervised learning is a useful approach for identifying fraud types that not previously identified. With this technique, the data examined for trends and anomalies that could be signs of fraud. These patterns used to categorize fresh transactions as fraudulent or lawful after they have discovered. (Bhattacharya, P., Bose, I., & Neogy, S. (2018). A few of the techniques of unsupervised learning are clustering and dimensionality reduction. Association Rule Learning, Anomaly Detection, Deep Generative Models:

2.5. Mobile Banking and Mobile Money Fraud in Ethiopia

This section explores the growing landscape of mobile banking and mobile money services in Ethiopia, alongside the associated challenges of fraud.

2.5.1. Mobile Banking and Mobile Money Evolution

The introduction of mobile technology has revolutionized the banking landscape, giving rise to mobile banking and mobile money services easy and accessible. These platforms offer customers convenient access to financial services through their mobile devices. While traditional banking channels like ATMs and internet banking have existed, mobile banking has significantly expanded accessibility and service delivery. For instance, Aijaz A. and Heikki (2015) highlight the substantial impact of mobile banking on the market, particularly in developing nations.

Mobile banking involves using mobile applications or USSD channels to perform banking transactions such as deposits, transfers, and balance inquiries. Mobile money, on the other hand, is a broader term encompassing mobile banking services along with additional features like person-to-person payments, bill payments, and mobile wallets. The emergence of platforms like M-Pesa in Kenya and M-Birr in Ethiopia has demonstrated the potential of mobile money to drive financial inclusion.

The origins of mobile money traced back down to the late 1990s and early on 2000s, when the thought of Mobile payments and digital wallets started to emerge (Anna, 2022). In 1997, Coca-Cola introduced hawking machines in Helsinki that allowed customers to buy in drinks through textual matter messages, this makes the first instance of a Mobile wallet and popularizing the concept of using Mobile devices for transactions. In 2003, Alibaba's Alipay launched, offering defrayment methods via cell phones. These milestones played a significant role in shaping the chronicle of Mobile money according to TechBullion's 2023 article “What Is a Mobile Wallet, Origin and History in Financial Technology?” Mobile money is a banking service that enables customer’s mobile phone to perform financial transaction using telecommunication networks. With Mobile money, subscribers can pay bills, receive money, and conduct business using virtual Mobile accounts well known as mobile money wallets. This serve allows users to swear directly from their phones without needing to visit a physical financial institution.(Peacock, 2021).

The first Mobile Money service in Ethiopia, M-BIRR, launched commercially in 2015 by MOSS ICT, a technology provider, in collaboration with six Ethiopian microfinance institutions. This

service offers mainly person-to-person money transfers, deposits and enables social payments to made by the Government to more than a million customers. Other services like HelloCash, Amole, and CBE Birr are also offered, and the majority of banks either currently offer Mobile Money services or are developing them (The Impact of Private Sector Projects in Africa - Studies from the EIB-GDN Programme, n.d.).

2.5.2. Mobile Banking and Mobile Money in Ethiopia

Ethiopia has witnessed rapid growth in mobile banking and mobile money adoption in recent years. Services offered by providers like M-Birr have expanded financial access to previously underserved populations. However, challenges such as network infrastructure, digital literacy, and security remain.

2.5.3. Fraud in Mobile Banking and Mobile Money

Types of Fraud

The increasing reliance on mobile banking and mobile money has made these platforms attractive targets for fraudsters. Common fraud types include account takeover, unauthorized transactions, SIM swapping, phishing, and social engineering. These fraudulent activities exploit vulnerabilities in the system and user behaviour illicitly acquire financial gains.

It is essential to note that the examples offered above are not an entire list of fraud types, patterns, or fraudulent acts, as other indicators or patterns that generate suspicion of illegal activity may exist. Financial institutions and organizations bear a significant responsibility for maintaining a state of constant vigilance in protecting customers from various attackers, and the same is true for the customers as well. This includes implementing strong and comprehensive monitoring systems that can detect suspicious transactions using many other technologies and processes as well which is required in the NBE fraud directive one or the other way.

Impact of Fraud

Fraudulent activities in the mobile banking and mobile money ecosystem have far-reaching consequences. Financial losses incurred by individuals and institutions, erosion of trust in digital financial services, and reputational damage are significant challenges. Additionally,

fraud can hinder financial inclusion efforts by discouraging users from adopting these services.

2.6. Related works

This section discuss the related works internationally and in an Ethiopian context. Few of the papers summarized in the table as local and international literatures in the following two tables in this section. Researchers worldwide are constantly working to improve detection methods, but fraudsters are equally adopt at developing new techniques to defeat detection by any means. This research analyses research papers and some Ethiopian-specific studies to identify key challenges and research gaps that need to addressed in the Ethiopian context.

A major challenge for Ethiopian banks lies in keeping pace with the ever-evolving tactics of fraudsters (Papers 1, 2, 7-see summary below). Additionally, the vast majority of financial transactions are legitimate, leading to imbalanced datasets that hinder machine-learning algorithms (Papers 4, 7). Furthermore, most research focuses on traditional bank fraud, neglecting the growing usabbility of mobile banking and mobile money services, which are particularly prevalent in Ethiopia (Mengash & Girma, 2021). Limited access to publicly available, labeled datasets for fraud detection further restricts research and development efforts in Ethiopia also being noted in both Ethiopian and international context.

Beyond these challenges, several research gaps require exploration in the Ethiopian context. The role of financial regulators in preventing and mitigating fraud needs further investigation, considering their potential impact on the effectiveness of existing control systems (Melese Gessese, 2022). While supervised learning techniques dominate fraud detection research, unsupervised anomaly detection methods hold promise for uncovering hidden patterns in fraudulent activities (Tewodros Yalew, 2021). The long-term effects of fraud on the Ethiopian financial sector, including bank stability and customer trust, are not well understood (Tewodros Yalew,2021). A deeper understanding of the root causes of fraud, such as ethical considerations and organizational culture, can inform preventative strategies specific to Ethiopia's banking environment (Techalu Setarge, 2022). Research is also needed to explore the effectiveness of existing preventative measures like employee training and improved monitoring (Techalu Setarge, 2022). Finally, fostering collaboration between banks,

regulators, and other stakeholders is crucial for combating fraud effectively a system wise detection is mandatory that run through every transaction.

Given the challenges of imbalanced datasets and the need for more sophisticated detection methods, the study noted that exploring the application of machine learning for fraud detection in Ethiopian banks is another exile to be explored in various contexts. Studies have shown that machine learning algorithms can be highly effective in identifying fraudulent transactions, even with imbalanced data (Al-Hashedi, 2021 [6]; Mytnyk et al., 2023 [8]). However, implementing machine-learning solutions requires careful consideration of several factors:

- **Data Availability:** As highlighted previously, the lack of publicly available, labeled datasets for fraud detection in Ethiopia is a hurdle. Banks can collaborate to create a shared data repository while ensuring data privacy maintained.
- **Technical Expertise:** Implementing and maintaining machine-learning models requires skilled data scientists and IT professionals. Universities and research institutions can play a role in building capacity in this area.
- **Regulatory Framework:** The regulatory environment around data privacy and algorithmic bias needs considered when deploying machine-learning solutions.

By addressing these challenges and research gaps, Ethiopian banks and researchers can develop more comprehensive and effective strategies to combat the evolving threat of financial fraud. One of the effective ways that a technology could be employed is using a Machine learning, when implemented thoughtfully and ethically, has the potential to be a powerful tool in this fight and his study will go through implementing a SVM as a fraud detection algorithm and evaluated in a real banking transition data in Ethiopian banking context. The following two tables (table 1 and table 2) summarizes the literatures selected for review from various sources presented.

Table 1 Fraud related papers in Ethiopian context

#	Author(s)	Year	Title	Key Challenges	Research Gap	Citation
1.	Tewodros Yalew	2021	Effects of Fraud on Bank Performance in Ethiopian Commercial Banks	The research focuses primarily on the immediate financial impact of fraud. Further studies could examine the long-term consequences of fraud on banks, including potential systemic risks or instability within the financial sector.	The research emphasizes the need for customer education on fraud risks. Further studies could explore the effectiveness of existing customer awareness, Conduct research on the impact of fraud across the entire Ethiopian financial sector, including microfinance institutions, savings & credit associations, and insurance companies.	Addis Ababa university College of business &economics Department of accounting and finance, unpublished
2.	Techalu Setarge	2022	FRAUD CAUSES AND EFFECT ON THE FINANCIAL PERFORMANCE: THE CASE OF ETHIOPIA BANKING INDUSTRY	While the study finds a negative impact of fraud on bank performance, performance scope is not detail considerations.	The study recommends training and improved monitoring, but a more in-depth analysis of effective detection and prevention strategies in the future with the help of technology	Department of accounting and finance College of business and economics Addis Ababa university
3.	Melese Gessese	2022	Evaluation of the role of NG|Screener Project on Fraud Prevention: The Case of Commercial Bank of Ethiopia	Internal Control System Weaknesses: Limited Coverage of technology used (NG| Screener in this case)	While the study recommends exploring machine learning and AI for fraud prevention, it does not delve into the specific functionalities or potential challenges of these technologies in the Ethiopian banking context.	Addis Ababa university College of business and economics School of commerce Department of project management
4.	Mengash, H., & Girma, A.	2021	An Assessment of Mobile Banking Security Challenges in Ethiopia.	- Weak authentication methods - Lack of robust security infrastructure - Limited customer awareness	- Need for research on context-specific security solutions for mobile banking in developing countries	Mengash & girma, 2021

No.	Title, Author, Publication Year	Descriptions	Key Findings Recommendations	Research Gap
1.	A systematic literature review on frauds in banking sector (Deepa Mngala and Lalita Soni, 2023)	Banking industry peculiarly has become soft target for several pernicious deceptive and fraudulent activities. The purpose of this paper is to systematically review the literature published in past 20 years on bank frauds and present a holistic view on causes and consequences of bank frauds and measures to curtail this menace.	• This paper provides a systematic review of bank fraud, highlighting the impact of control environment loopholes on fraud occurrence and the deleterious effects on stakeholders. It also discusses Anti-fraud measures such as internal auditing; machine learning, fraud detection, and forensic accounting adopted by banks. The study emphasizes the need for a threefold approach of deterrence, prevention, and awareness to curb fraudulent activities. • The study emphasizes the need for multiple fraud prevention and detection techniques in the banking sector due to the evolving nature of fraud methods.	• It also highlights the importance of studying the role of financial regulators in combating bank fraud, as this area has not been extensively explored in existing literature. • Therefore, studies in this area offer valuable insights into various aspects of fraud and its prevention. • Financial regulators play a critical role in the fight against financial crimes, which is not much studied in the existing literature. Thus, future research may explore the role of financial regulators in curbing bank fraud.
2.	INVESTIGATION INTO THE RISKS FACING MOBILE BANKING: A CASE OF COMMERCIAL	The objective of the study was to investigate the risks facing mobile banking among the Commercial banks in Kenya.	Risks for commercial banks in Kenya due to mobile banking:	There is a need to undertake further studies to establish the effects of the risks established on the profitability of commercial banks in Kenya as the mobile banking offer hit by the fraud

	BANKS IN KENYA (JOHN NJAU KARANJA, 2017)		• Unauthorized individuals penetrating web servers to manipulate information. • Intruders accessing the organization's emails. • Cybercriminals transferring organization's secrets to software that then transmits them to the open Internet. • Hackers stealing mobile banking PINs and codes. • Developers breaching the system to transfer customer funds.	happenings because of lack of enough security measures in place.
3.	Financial Fraud Detection using Machine Learning Techniques (Matar Al Marri, Ahmad AlAli, May 2020)	• With the advent of AI, ML-based approaches used to detect fraudulent transactions, and the Kitchenham methodology approach used for SLR analysis. • -Popular ML techniques used for fraud detection, common fraud types, and evaluation metrics summarized. • -identified unexplored or less studied algorithms for fraud detection.		• Most of the research focuses on supervised learning but also on bank fraud (credit card and bank statement fraud), insurance fraud, mortgage fraud, and health fraud. Engage in the same ML techniques, but on mobile banking and mobile money services in a local context. • future research to go through unsupervised learning on mobile banking and mobile money services,

		• -Financial fraud detection is focused on supervised learning in most research. • -Ensemble methods that take advantage of multiple algorithms have been a rising trend recently. • -Unsupervised learning approaches were less commonly employed. • -Future research recommended to go through unsupervised learning, like anomaly detection, can cover new insights.		
4.	Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review (Ali, A.; Abd Razak, S.; Othman, S.H.; Eisa, T.A.E.; Al-Dhaqm, A.; Nasser, M.; Elhassan, T.; Elshafie, H.; Saif, A.	Conventional techniques such as manual verifications and inspections are imprecise, costly, and time consuming for identifying such fraudulent activities. With the advent of artificial intelligence, machine-learning-based approaches can be used intelligently to detect fraudulent transactions by	The research identified five main categories of financial fraud addressed using machine learning (ML) techniques: 1. **Credit Card Fraud:** 2. **Financial Statement Fraud:** 3. **Insurance Fraud** 4. **Financial Cyber-Fraud:**	• With the advent of AI, ML-based approaches can be used to detect fraudulent transactions, and the Kitchenham methodology approach is used for SLR analysis in the study. • Financial fraud detection is focused on supervised learning in most researches identified but on the case of the fraud types on finical,

	Published: 26 September 2022)	analysing a large number of financial data. Therefore, this paper attempts to present a systematic literature review (SLR) that systematically reviews and synthesizes the existing literature on machine learning (ML)-based fraud detection.	5. **Other Financial Fraudulent Types:** This likely encompasses a broader range of less common or specialized fraud activities.
			
			The research found a wide variety of ML algorithms used for fraud detection, with some of the most common ones being:
			
			• **Supervised Learning:**
			  ○ SVM, HMM: Hidden Markov Models, ANN, Fuzzy Logic, KNN, Decision Tree, Logistic Regression,
			• **Unsupervised Learning:**
			  ○ Clustering:
			• **Approaches:**
			  ○ **Ensemble Methods:** Combining multiple algorithms for improved accuracy and robustness.
			  ○ **Random Forest:** Combining multiple decision trees for high accuracy.
			  ○ **Naïve Bayes:** Probabilistic algorithm for efficient fraud classification.
				credit card and the like however the mobile banking and mobile money has not been touched using these methods.
				• Future research recommended to go through unsupervised learning, like anomaly detection, can cover new insights.
				• An important point to consider further is that while most research focuses on supervised learning for detecting fraud in areas such as bank fraud (credit card and bank statement fraud), insurance fraud, mortgage fraud, and health fraud, there is a lack of research on fraud detection using the same machine learning techniques for mobile banking and mobile money services in a local context.

6.	Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019 (Khaled Gubran Al-Hashedi , Pritheega Magalingam15 April 2021)	This paper gives a comprehensive revision of the state-of-the-art research in detecting financial fraud from 2009 to 2019 inclusive and classifying them based on their types of fraud and data mining technology utilized in detecting financial fraud.	• Few numbers of papers have been published on financial fraud detection in the past decade and it was noted that most of the researchers categorized fraud types into three main groups such as Bank, Insurance, Financial statement fraud.	•
7.	Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances (Waleed Hilal a, S. Andrew Gadsden a,*, John Yawney, 2022)	The significance of detecting fraud and its detrimental effects on the financial economy was highlighted in this paper, along with the associated challenges of applying anomaly detection techniques to combat this continually growing problem. The main areas explored in this survey were credit card fraud, insurance fraud and money laundering. It was made evident that the challenges faced varied significantly based on the different fraud applications	• Different fraud types pose unique challenges (real-time detection crucial for credit card fraud). • No way of universal anomaly detection technique works for all fraud types.	• Lack of publicly available labelled datasets hinders research and development. • Imbalanced datasets due to rare fraudulent cases since data labelled as fraud is less in the data set and the method was many implemented on the card banking data set but not the mobile banking or mobile money data sets.
8.	Application of Artificial Intelligence for Fraudulent Banking Operations Recognition Bohdan Mytnyk, Oleksandr Tkachyk ,	The article discusses the increasing prevalence of financial fraud in the digital era and the role of artificial intelligence (AI) and machine learning (ML) in	It proposes various classification algorithms, particularly an artificial neural network model, which significantly improves accuracy in identifying fraudulent transactions. The study also presents methods	This study focuses only on identifying fraudulent transactions in online banking, and different detection methods may be needed for other types of financial fraud.

	Nataliya Shakhovska , Solomiia Fedushko and Yuriy Syerov, Published: 10 May 2023)	detecting and preventing such frauds	for enhancing detection accuracy, such as managing imbalanced datasets, feature transformation, and feature engineering. The results show that all selected algorithms perform well in recognizing fraudulent bank transactions, with logistic regression algorithm performing the best. The study's application of artificial intelligence to identify fraudulent banking transactions is particularly relevant in the current context, with increased online transactions during the pandemic and heightened charitable activities during times of conflict.	

Table 2 Few of International papers on Fraud related topics summrized

Additional Research Gaps

According to Ali, A et al. (2022) articles relating to financial fraud detection using ML approaches from 2010 to 2021 is shown in Figure 1 below, time line against the number of articles.

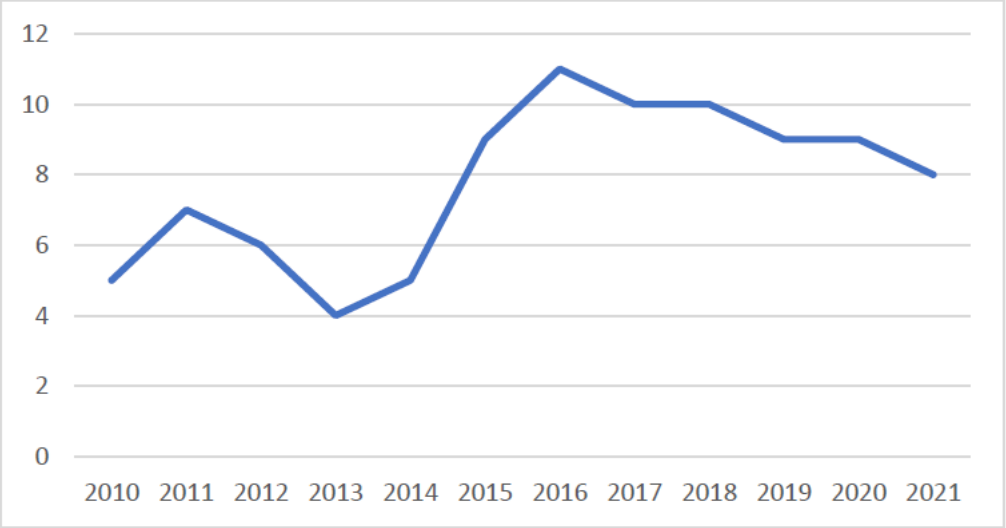


Figure 1 Summary of articles in a year. (Adopted from the Ali .A 2022)

On this same article, the literature's reviewed, it is understood that all of the articles are all about credit card fraud, financial statement fraud, insurance fraud, and financial cyber fraud, which are the main research articles that the fraud detection using through machine learning applications.

The figures presented above from research by Ali A. et al. in 2022 highlight a shortage of studies globally on financial payment fraud, especially concerning mobile banking and/or mobile money transactions. This shows a very significant research gap that requires attention, as discussed in the problem statement within this study. Therefore, this study aims to explore and test specific machine learning techniques to contribute this critical research gap. By focusing on the context of Ethiopian banks, this research underscores the importance of effective fraud detection strategies in enhancing fraud management practices in the technological aspect. The findings will contribute to the existing literature and provide valuable insights into mitigating fraud in mobile banking environments, paving the way for future research and practical applications in this field.

Fraud Type	Description	Technique Used	No. of Reference
Financial Statement Fraud	This is a corporate fraud such that the financial statements are illegitimately modified to allow the organizations to look more beneficial.	Support Vector Machine	20
		Clustering based method	
		Decision Tree	
		Logistic Regression	
		Naïve Bayes	
		Artificial Neural Network	
Fraud Type	**Description**	**Technique Used**	**No. of Reference**
Credit Card Fraud	Illegitimate use of the card without proper owners' authorization	Support Vector Machine	32
		Fuzzy logic	
		Clustering based method,	
		Artificial Neural Network	
		Hidden Markov model	
		Decision Tree	
		Genetic Algorithm	
		Artificial Neural Network	
		Naïve Bayes	
		Logistic Regression	
		Random Forest	
Health Insurance Fraud	Fraudulent claims by individuals or organizations to support the relevant expenses of theft or accidental damages.	Support Vector Machine	5
		Artificial Neural Network	
		K Nearest Neighbors	
		Naïve Bayes	
		Clustering-based method,	
Auto Insurance Fraud	Fraudulent claims by an individual to get health insurance profits.	Support Vector Machine	3
		K Nearest Neighbors	
Cyber Financial fraud	Financial fraudulent activities through cyber space	Artificial Neural Network	3
		SVM	2
Others	Other frauds that are faced in the financial domains include commodities and securities fraud [32], mortgage fraud, corporate fraud, and money laundering.	Support Vector Machine	5
		Decision Tree	
		Fuzzy logic	
		Clustering-based method,	
		Hidden Markov model	

Figure 2 different types of financial fraud with the number of researches in each of the ML techniques (Adopted from the Ali .A 2022)

CHAPTER THREE

3. Research Design and Methodology for fraud detection on mobile banking and mobile money

In this chapter, the research design, methodology, and frameworks used in this specific study on mobile banking and mobile money services. The objective is to identify fraudulent transactions and propose effective fraud detection measures using machine learning, with a specific focus on Support Vector Machines (SVM). The study aims to minimize the negative impacts of fraud on customers and employees in the banking sector.

3.1. Research Design

This study employs a supervised machine learning approach to develop and evaluate a model for detecting fraudulent transactions in the Ethiopian banking sector. The CRISP-DM (Cross-Industry Standard Process for Data Mining) framework adopted to ensure a structured and repeatable process for model development, covering data understanding, preparation, modelling, and evaluation. CRISP-DM six key phases, each playing a critical role in building and deploying the model in this study as following section. This study employs a quantitative research method, utilizing machine learning and an experimental research design.

3.2. CRISP-DM Phases

3.2.1. Business Understanding and Data Acquisition

1. **Business Understanding:** In collaboration with banks experts is this research tries to understand their specific mobile banking and mobile money services fraud challenges. In addition, the research will try identifying key performance indicators (KPIs) for evaluating the model's effectiveness, such as the successful detection of fraudulent transactions. Additionally, the researcher will try to understand their risk tolerance and desired outcomes, tailoring the model's sensitivity accordingly.
2. **Data Acquisition:** Working with the chosen financial institution, Data will be captured in csv format from the database sources, which will be explained in the next chapter, which

the data is a comprehensive historical dataset of mobile banking and mobile money services transactions. This data will encompass details like transaction amounts, timestamps, merchant information, device specifications like USSD android or other , and user behaviour patterns in terms of transaction behaviour

3.2.2. Data Pre-processing and Feature Selection (CRISP-DM Phases 3 & 4):

3. **Data Pre-processing:** To ensure model accuracy and consistency, the data will undergo thorough cleaning. This may involve handling missing values (e.g., imputing missing data with user's typical assumptions), correcting inconsistencies (e.g., standardizing date formats), and removing outliers that could skew the model's learning (e.g., unusually high transactions).
4. **Feature Selection:** the study explore the data to identify features most relevant for fraud detection. This might involve creating new features based on existing ones. For example, a new feature could represent the ratio of a transaction amount to the user's average daily spending. Additionally, time-based features could be created by comparing transaction times to a user's typical spending patterns (e.g., late-night transactions outside of usual spending hours).

3.2.3. Model Selection and Training (CRISP-DM Phase 5):

5. **Model Selection:** Considering the characteristics of the data and desired model performance, model chosen a suitable supervised machine-learning algorithm. Common options include Logistic Regression, Random Forest, Support Vector Machines (SVM), or Gradient Boosting. Each algorithm offers advantages and disadvantages, the choice is based on factors like interpretability, accuracy, and computational efficiency as well has capacity of ingesting and responding to large data set, which SVM is best at it.

Model Training: The selected machine-learning algorithm will undergo training using pre-processed data, enabling it to learn from historical transactions categorized as either fraudulent or legitimate. Training involves feeding the data into the model, enabling it to identify patterns and relationships between features and their corresponding labels.

3.2.4. Model Evaluation and Deployment (CRISP-DM Phase 6):

6. **Model Evaluation:** The model's performance will be assessed on a separate holdout dataset that was not used during training. Metrics such as accuracy, precision, recall will be employed to evaluate how well the model generalizes to unseen data and its effectiveness in detecting fraudulent transactions. Based on the evaluation results, the model parameters may be refined, feature selection adjusted, or alternative algorithms considered if necessary.

Deployment and Monitoring: Once a satisfactory model achieved, the result reported and presented for manual evaluation and review by the fraud investigators opinion and by far, it can be integrated with the mobile banking and mobile money infrastructure for real time detection. The model will score incoming transactions in real-time, flagging potentially fraudulent activities for further investigation or triggering automated security measures but it is out of scope in this research project. Continuous monitoring of the model's performance is crucial. As fraudsters evolve their tactics, the model may need retraining or adjustments to maintain its effectiveness over time. This may involve incorporating new data points or adapting to new types of fraudulent activities to apply real life.

This study employs a quantitative research method, utilizing machine learning and an experimental research design. The CRISP-DM framework will guide the development process, ensuring a robust and efficient machine-learning model for fraud detection in mobile banking and mobile money transactions.

3.2.5. Methodological Framework Using SVM for Fraud Detection

Selecting SVMs for Mobile Banking Fraud Detection

Support Vector Machines (SVMs) are a strong choice for fraud detection in mobile banking and mobile money services due to several key advantages. First, SVMs are well suited for handling high-dimensional data, a common characteristic in mobile banking transactions, which often include features beyond transaction amounts, such as timestamps, locations, device information, and anonymized user details (James et al., 2021). By identifying optimal

hyperplanes, SVMs effectively separate fraudulent and legitimate transactions in complex, high-dimensional spaces (Cristianini & Shawe-Taylor, 2000).

However, a notable challenge in fraud detection is class imbalance, where fraudulent transactions are far less frequent than legitimate ones (Ahmad et al., 2023). This imbalance can cause standard SVMs to bias towards the majority class (legitimate transactions), reducing the model's ability to detect fraud. To address this, techniques like class weight adjustment or cost-sensitive SVMs can be employed to improve the detection of minority class instances—fraudulent transactions (Japkowicz & Stephen, 2002).

Another key advantage of SVMs is their interpretability compared to other machine learning models such as deep neural networks. While deep learning models can achieve high accuracy, they require large labelled datasets, which are often unavailable in fraud detection contexts, and are computationally intensive to train (Montaño & Garcia-Chamizo, 2018). Moreover, deep learning models typically lack transparency, making it difficult to explain why a particular transaction flagged as fraud. In contrast, SVMs provide clearer decision boundaries, which are crucial in fraud detection applications where explainability is necessary for understanding and justifying detection outcomes.

Overall, SVMs offer a compelling combination of effectiveness in high-dimensional data, robustness in handling fraud detection, and relative interpretability. These characteristics make SVMs a strong contender for fraud detection in mobile banking and mobile money services in this study.

In conclusion, SVMs offer a compelling combination of effectiveness in high-dimensional data, robustness in handling fraud detection problems, and relative interpretability. These characteristics make them a strong contender for supervised learning algorithms in this study for fraud detection application within the mobile banking and mobile money domains

3.2.6. Data Collection and Sampling

Data Sources

Transactional and user related data will be collected from bank with specific Ethiopian bank will be established to obtain data from their mobile banking and mobile money platforms, which is already stated in the scope section. The data will include details like transaction amounts, timestamps, locations, and potentially user information (with anonymization techniques following data privacy regulations) (Memon et al., 2022).

In this study, data collected from the banking databases of mobile banking and mobile money from the bank databases for analysis, which have transaction data of one month for both. This data is just private data from customers and banks, which requires high security or masking to protect customer data utilized for another purpose. In this study, the data kept secure, and the data encrypted and destroyed for security reasons after the study completed. Another challenge for having this large data set could be a machine capable of executing such a large data size for the researcher in this study.

While the data quality has generally been good, a preliminary review revealed scattered missing data points. These gaps could potentially introduce bias or skew the results of the analysis if not addressed effectively. To ensure a reliable and representative sample, we will implement the following measures:

- Identifying the extent of missing data: systematically analysed the dataset to determine the number and distribution of missing points across variables and observations.
- Investigating the reasons for missing data: Understanding why certain data is missing (e.g., technical errors, non-response, incomplete records) is crucial for determining the appropriate imputation strategy. In this study therefore, the researcher consulting the business owners on the situation and they respond as the cases is could be error, transaction failure because of various reasons like network interruption and some other unknown reasons for the values to missed.

Sampling Approach

Due to limitations in data access, a convenience sampling approach is used. This method involves selecting readily available transaction data, with random sampling within specific periods to avoid bias. Data from one week in December 2023 and from select days across five months (July to November 2023) are included in the analysis, ensuring a diverse range of transactions represented.

3.2.6.1. Source of data, population and Sampling

This section outlines the data source, target population, and sampling approach utilized in this study. Due to data accessibility limitations, the study focused on recent transactions (i.e. use most of the data from the last week of December). The convenience sampling approach is used but has may also introduces a bias, it offers the advantage of capturing evolving fraud patterns.

This research employs quantitative analytics to examine the technical aspects of fraud detection. Financial transaction data and associated user audit information, anonymized for security and ethical reasons obtained from the bank. The focus was on analysing transactions flagged for potential fraud and those confirmed as fraudulent using selected feature.

The rationale behind why quantitative analysis is chosen is because quantitative analytics allows for the identification of patterns, anomalies, and statistical indicators within the dataset that might be indicative of fraudulent activity [James et al., 2021]. This involves techniques such as data mining, statistical analysis, and data visualization to extract valuable insights from the data [James et al., 2021].

The study sought to leverage the influence of quantitative analytics to enhance fraud detection capabilities research domain stated. By analysing the technical aspects of these transactions and considering user audit information, this study aimed to develop effective strategies and methodologies for detecting and preventing fraudulent activities in the digital financial landscape. Following this, the next paragraphs will present details of the data source and related activities in the study.

Data Sources

This study utilizes data from internal sources, specifically the transaction and user audit log information stored within the mobile banking and mobile money systems of the bank. These datasets offer valuable insights into user behaviour and transaction patterns, but they require cleaning and pre-processing before analysis.

- **Target Population:** The target population of the study under consideration is the customer's transactional data for mobile banking and mobile money services. Due to limitations in data accessibility, this study utilizes a convenience sample extracted from the target population. This sample consists of six months of transaction data:

    - **Dataset 1:** A one-week sample (December 23rd - 30th, 2023) focused on recent transactions and includes data flagged as suspicious or fraudulent by the bank's fraud management system. (Both fraudulent and all transactions taken)
    - **Dataset 2:** A five-month sample (one day from each week in July, August, September, October, and November 2023) containing a broader range of mobile banking and mobile money transactions.

- **Accessible Population:** the data to be under analytics could only be one-month data of both the mobile banking and the mobile money services.

The transaction per day for each of the cases were in reaches to millions, which require intensive hardware machines to run various computation to find results. However, for this study the researcher takes six month of data form the mobile banking and mobile money transaction form the database schema based on the convenience sampling technique explained.

Sampling Rationale

A convenience sampling approach, a non-probability sampling technique where participants or data are readily accessible (Golzar, 2022), is employed for this study. The rationale for using a convenience sample a non-probability sampling technique where researcher select participants or data readily available (Golzar, 2022). This due to several factors as follws:

- **Accessibility:** The research is limited to data accessible from the bank.

- **Computational Resource Requirement**: Requirements of computational resources to consider large data set than this one is impossible.
- **Timeliness:** Including recent transaction data (December 2023) allows for analysis of the most current fraud patterns.

Data Selection and Randomness

While the overall sampling approach is convenience-based, the specific dates within datasets 1 and 2 randomly chosen to avoid bias towards any particular period within the six-month timeframe.

Data Size and Considerations

Tables 1 and 2 below present the number of transactions included in each dataset. The large volume of transactions in Dataset 1 (December 23rd - 30th, 2023) highlights the potential need for computational resources during analysis. This has been addressed by considering appropriate hardware and software capability to ensure efficient processing.

Limitations

The use of a convenience sample acknowledges the limitation that the findings may not be generalizable to the entire population of mobile banking and mobile money transactions. However, the inclusion of a recent dataset (December 2023) can offer valuable insights into current fraudulent activities hence it works well on model training and development process.

Mobile Banking data set summary:

For analysis, the dataset with a sample of mobile banking was composed of completed, cancelled and paid transactions inclusive all transactional information without any cleaning. Date is randomly selected using convenience sampling and summary of count presented below.

Table 3 Randomly selected six months of data from mobile banking transactions.

Date	# Transactions (including complete, paid and cancelled)
July **01,2023**	2,055,707
August **07, 2023**	2,148,759
September **15, 2023**	1,984,315
October **24, 2023**	1,945,927
November **29,2023**	2,201,017
December **23-31,2023**	17,725,626
Total number of transactions	24,131,109

Mobile money data set Summary:

The sample data set for mobile banking collected for analysis includes completed, cancelled, and paid transactions, along with all transaction information without any cleaning. Transaction of the date randomly selected.

Table 4 Randomly selected six months of data from mobile money transactions.

Date	# Transactions (including complete, paid and cancelled)
July **20,2023**	144,973
August 16 **, 2023**	142,408
September **11, 2023**	161782
October **21, 2023**	205,403
November **14,2023**	89,141
December **23-31,2023**	1,234,493
Total number of transactions	1,978,200

CHAPTER FOUR

4. Data Collection, Analysis, and Discussion

Analysis using CRISP-DM:

4.1. Business understanding:

To understand the operations of mobile banking and mobile money services, the researcher conducted brainstorming sessions with business owners in banks. These sessions aimed to gain a deeper insight into both products. This research paper includes various observations, covering formal and informal discussions, along with occasional incidents of fraudulent activities that occurred earlier. According to the business owners customers sometimes even bank employees share information lead have been tricked mainly by social engineering techniques such as phishing through phones, mishandling secret information like PINs and passwords, and rarely hacking customer accounts on both customer and banking systems. The notes taken presented below:

4.1.1. Mobile banking services:

This service is much like the customer holds his bank account and the banking service in his pocket though a mobile cell phone or tablet, which has two options using:

- USSD (Unstructured Supplementary Service Data) is a technology used by mobile network operators to provide interactive services to mobile phone users. It allows users to access various services and perform transactions by dialling specific codes on their mobile devices. USSD codes usually start with a "*" (asterisk) symbol followed by a sequence of numbers and end with a "#" (hash) symbol.
- An application developed by the bank on an Android or iPhone device. In this scenario, the bank authenticates the application to prevent fake apps and utilizes an activation code for one-time authentication. Additionally, the mobile banking application utilizes a Wi-Fi or data network to provide services.

4.1.2. Mobile money services:

Mobile money services in Ethiopia include popular platforms like CBE Birr and Tele Birr. These services differ from traditional banking in that they do not require a bank account and instead operate solely through mobile devices.

Unique Service Features:

- Customer Identification: Their phone numbers uniquely identifies customers, and the service is tied to the SIM card rather than a bank account.
- Transaction Methods: Mobile money allows transfers between mobile wallets or to a customer's bank account. Deposits are stored on the mobile device, and all transactions are conducted through USSD or mobile applications.

These services align with the National Bank of Ethiopia's agenda to promote financial inclusion and digitization, as outlined in their three-year strategic plan.

4.2. Data collection:

The purpose of data collection in this study was to obtain transactional data from mobile banking and mobile money platforms that could be used to train and test an SVM-based fraud detection model. The data needed to encompass a wide range of transaction types, including both legitimate and fraudulent activities, to ensure that the model could effectively distinguish between normal and suspicious behaviour. Additionally, the data was required to contain relevant features, such as transaction amounts, timestamps, and customer identifiers, which are critical for the accurate classification of transactions.

4.2.1. Sources of Data

The data used in this research sourced from the bank's RDBMS (Relational Database Management System). Specifically, the training dataset consisted of fraud-classified data provided by audit and fraud management teams. The testing data derived from labeled fraud data. All reported fraud cases linked to mobile banking/mobile money systems.

Two separate datasets collected during the analysis phase:

- **Mobile Banking Dataset:** Consists of 3,648,970 transactions, including both successful and cancelled transactions. These transactions are associated with mobile banking customers and often involve suspicious phone numbers, some of which linked to mobile money services or reported instances of fraud.
- **Mobile Money Dataset:** Includes 1 million transactions from the mobile money platform, out of the 25 million registered users. This dataset contains detailed information on various aspects of mobile money services.

Details of the data fields presented below.

Table 5 Mobile banking transactions pre-processed data.

#	Data column	Description
1	DB_ID	This field holds database identifier or database ID which is a unique identifier assigned to a specific database within a database management system (DBMS) or a database instance.
2	AMOUNT	Transaction amount but holds 0 in the database.
3	SALES_TAX	This field hold the transaction sales tax but hold all the same value of 0
4	TAX_CODE	This field holds the transition code number but holds a text data “ZERO”
5	CHARGE_CODE	This field holds charge codes for a particular type transition but hold no data in it
6	PROMO_IDS	This field holds stores promotional or discount code identifiers. It is a way to keep track of the unique identifiers assigned to different promotions or discounts within the database.
7	MOBILE_APP_ID	This field explains the payment method for the specific transaction, and the customer has three choices in this regard (Android app, USSD and iPhone)
8	MIN_AMOUNT	This is a field to holds a data about the promotion id min payment amount but now has no data
9	BASKET_ID	A field that stores a unique identifier for a shopping basket or cart. It is used to associate specific items or products with a particular shopping basket or cart within the database.
10	CHANNEL_NAME	Has only a text data “Mobile”
11	CURRENCY	Has only data “ETB”
12	CUSTOMER_ID	This field holds a customer unique identifier for mobile banking data

#	Data column	Description
13	ID	A unique identifier of the account holders
14	PAYMENT_METHOD_ID	This filed holds the debit account id and additional account ID
15	PAYMENT_METHOD_DESCRIPTION	This field hold the debit side customer name
16	PAYMENT_METHOD_NUMERIC	This field holds payment method unique identifier
17	PAYMENT_METHOD_TITLE	The data field for this holds as the PAYMENT_METHOD_DESCRIPTION
18	PROCESSED_DATE	Holds time stamp of the transactions processed
19	VERB	NO data
20	PRODUCT_NAME	This filed holds the sms message send to customers the explains about the transaction like sender and receiver names transaction reference number i.e. “ETB 100.00 debited from ES[redacted]U A[redacted]A GO[redacted]O-ETB-5528 for Topup 092[redacted]77 (Topup 092[redacted]77) on 30-Jan-2024 with transaction ID: FT240301VRQ3.” *For security reasons few of the information displayed here is shaded in black*
21	RECIPIENT_ID	The field contains the customer ID of the credited customer.
22	VENDOR_ID	This filed holds the debit customer account with mobile banking customer ID like “ FF5AC8C6CE55EC1-1-100[redacted]31”
23	PRODUCT_ID	This filed holds customer credit account customer account number
24	RECIPIENT_NAME	This field holds credit account user name
25	AUTH_CODE	This field holds no data
26	AUTH_CODE_REQUIRED	This field holds 0 as data
27	RFID_REQUIRED	This field holds no data
28	AND_FUNCTION	This field holds 0 as data
29	AUTH_CODE_EXPIRY_DATE	This field contains the transaction expiry time in seconds
30	PRICE_OVERRIDE	Has 0 as a data
31	PRICE_CASCADE	Has 0 as a data
32	STATUS_CODE	This filed holds for the transaction processing status as completed, cancel … etc.
33	TRANSACTION_TYPE	This field holds for the transaction type ad debit
34	ENCRYPTION_SEED	This field typically refers to a specific field or column within a database table that stores the encryption seed or key for encrypting and decrypting data but has a record “NONE”
35	INITIATE_FULFILLMENT	This field stores information related to the initiation or status of fulfilment processes for orders or requests has just 1 as a data
36	DELIVER_MOBILE	This field hold 0 only as a data
37	DELIVER_EMAIL	Has has data
38	DELIVER_ADDRESS	Has no data

#	Data column	Description
39	SELECTED_OPTIONS	This filed holds the transaction amount and the reason that is feed by the customer. i.e. “param.decimal.Amount: 6000; Reason: HA; ““param.decimal.Recharged Mob No: 099[redacted]31; param.decimal.Amount: 10; “
40	RFID	Has no data
41	RECEIPT_ART_WORK_ID	Has no data
42	USE_SEED	Hold 1 and 0 as data
43	CREATION_TIMESTAMP	This field stores the timestamp or date when a record was initially created or inserted into the table. It helps to track when the data was first added to the database.
44	MODIFICATION_TIMESTAMP	This field captures the timestamp or date of the most recent modification made to a record in the table. It is updated whenever any changes are made to the data, allowing for tracking the last time the record was modified.
45	VERSION	The "VERSION" field is typically used to track the version or revision of a record within a table. It is often implemented as an incremental number or a timestamp-based value. Each time a record is modified, the version number is incremented or updated, indicating that a new version of the record has been created.

Table 6 Mobile Money service of the customer transaction data Colum description.

#	Data column	Description
1	ORDERID	This field stores a unique identifier for transaction, typically used to identify and reference a specific order within database.
2	TRANS_STATUS	This field holds the status or state of a transaction. It may contain values such as "Completed," "authorized," "cancelled," etc. or other relevant statuses indicating the current state of the transaction.
3	TRANS_INITIATE_TIME	This field captures the timestamp or date when the transaction was initiated.
4	TRANS_END_TIME	This field records the timestamp or date when the transaction was completed or ended.
5	DEBIT_PARTY_ID	This field holds the identifier or reference for the party or entity that is debited in the transaction.
6	DEBIT_PARTY_TYPE	This field indicates the type or category of the debit party, such as "1000," "5000," or any other classification.
7	DEBIT_PARTY_ACCOUNT	This field contains the account number or identifier associated with the debit party.

#	Data column	Description
8	DEBIT_ACCOUNT_TYPE	This field represents the type of account associated with the debit party, such as "10009," "21205," or any other account type
9	DEBIT_PARTY_MNEMONIC	This field store a mnemonic or shorthand identifier for the debit party, which holds phone number or a short code for company data and the name…
10	CREDIT_PARTY_ID	Similar to DEBIT_PARTY_ID, this field stores the identifier or reference for the customer who is credited in the transaction.
11	CREDIT_PARTY_TYPE	This field represents the type or category of the credit party, similar to DEBIT_PARTY_TYPE.
12	CREDIT_PARTY_ACCOUNT	This field contains the account number or identifier associated with the credit party.
13	CREDIT_ACCOUNT_TYPE	This field represents the type of account associated with the credit party, similar to DEBIT_ACCOUNT_TYPE.
14	CREDIT_PARTY_MNEMONIC	This field may store a mnemonic or shorthand identifier for the credit party, similar to DEBIT_PARTY_MNEMONIC.
15	EXPIRED_TIME	This field stores the timestamp or date when a transaction or request is set to expire or become invalid if not completed within a specific timeframe.
16	REQUEST_AMOUNT	This field represents the requested amount for the transaction.
17	REQUEST_CURRENCY	This field indicates the currency in which the request amount is specified which is all in ETB
18	EXCHANGE_RATE	If applicable, this field stores the exchange rate used for converting between different currencies but data contains only “1”
19	ORG_AMOUNT	This field represents the original amount associated with the transaction, before any conversions or adjustments but holds dame as ACTUAL_AMOUNT.
20	ACTUAL_AMOUNT	This field stores the actual amount involved in the transaction, same as ORG_AMOUNT
21	FEE	This field represents any fees associated with utility payments like telecom, electricity
22	COMMISSION	This field stores the commission amount, if any, associated with the transaction but no data in the current database.
23	TAX	This field indicates any applicable taxes associated with the transaction but no value found.
24	ACCOUNT_UNIT_TYPE	This field represents the unit type or denomination of the account, such as "1001" for ETB
25	CURRENCY	This field indicates the currency of the transaction has only ETB

#	Data column	Description
26	IS_REVERSED	This field is an indicator (e.g., 0 or 1) that represents whether the transaction has been reversed or not.
27	REMARK	This field allows for additional comments or notes related to the transaction has various text fields.
28	IS_PARTIAL_REVERSED	This field indicates whether a partial reversal of the transaction has occurred but no data in it.
29	IS_REVERSING	This field is a binary indicator that represents whether the transaction is in the process of being reversed.
30	CHECKER_ID	This field stores the identifier or reference of the user or entity responsible for checking or approving the transaction.
31	REASON_TYPE	This field represents category associated with the transaction, such as "10002562," "1000043," or any other relevant classification.
32	LAST_UPDATED_TIME	This field captures the timestamp or date when the transaction record was last updated.
33	VERSION	This field represents the version or revision of the transaction record but no records in this filed exist.
34	LOAD_DATA_TS	This field stores the timestamp or date when the transaction data was loaded into the system.
35	ACCUMULATOR_UPDATE	This field stores information related to accumulators or aggregated values that are updated as a result of the transaction but has null record.
36	ACCUMULATOR_REVERSAL	This field indicates whether the transaction includes a reversal of accumulated values has null record has null record.
37	CHG_RATING_DETAILS	This field may store information related to rating or pricing details associated with the transaction has null record.
38	BANK_CARD_ID	If applicable, this field stores the identifier or reference for a bank card associated with the transaction has null record has null record...
39	BANK_ACCOUNT_NUMBER	This field Apologies, but I seem to have hit the response character limit. Here's the continuation
39	BANK_ACCOUNT_NUMBER	This field stores the bank account number associated with the transaction has null record.
40	BANK_ACCOUNT_NAME	This field contains the name of the bank account associated with the transaction has null record.
41	FI_ACCOUNT_INFO	This field may store additional information related to the financial institution or bank account involved in the transaction has null record.
42	DISCOUNT_AMOUNT	This field represents any discount applied to the transaction amount has only 0 as a record.

#	Data column	Description
43	REDEEMED_POINT_TYPE	If applicable, this field stores the type or category of redeemed points associated with the transaction has null record.
44	REDEEMED_POINT_AMOUNT	This field indicates the amount of redeemed points used in the transaction has null record.
45	IS_MAIN	This field is a binary indicator that represents whether the transaction is considered the primary or main transaction within a set of related transactions has null record..
46	CONSUMED_BUNDLE	This field refers to a consumed bundle or package associated with the transaction has null record.

	DB_ID	AMOUNT	SALES_TAX	CHARGE_CODE	PROMO_IDS	MIN_AMOUNT	AUTH_CODE	RFID_REQUIRED	INITIATE_FULFILLMENT	DELIVER_EMAIL	DELIVER_AD
count	3.608750e+06	3608750.0	3608750.0	0.0	0.0	3.608749e+06	3.000000	3.608747e+06	3.608746e+06	3608744.0	360
mean	1.261245e+18	0.0	0.0	NaN	NaN	2.772129e+05	0.333333	2.771045e-07	1.000000e+00	0.0	
std	4.999962e+16	0.0	0.0	NaN	NaN	5.266132e+08	0.577350	5.264072e-04	7.444523e-04	0.0	
min	1.211050e+18	0.0	0.0	NaN	NaN	0.000000e+00	0.000000	0.000000e+00	0.000000e+00	0.0	
25%	1.211050e+18	0.0	0.0	NaN	NaN	0.000000e+00	0.000000	0.000000e+00	1.000000e+00	0.0	
50%	1.311050e+18	0.0	0.0	NaN	NaN	0.000000e+00	0.000000	0.000000e+00	1.000000e+00	0.0	
75%	1.311050e+18	0.0	0.0	NaN	NaN	0.000000e+00	0.500000	0.000000e+00	1.000000e+00	0.0	
max	1.311050e+18	0.0	0.0	NaN	NaN	1.000392e+12	1.000000	1.000000e+00	2.000000e+00	0.0	

	DB_ID	AMOUNT	SALES_TAX	CHARGE_CODE	PROMO_IDS	MIN_AMOUNT	AUTH_CODE	RFID_REQUIRED	INITIATE_FULFILLMENT	DELIVER_EMAIL	DELIVER_AD
count	3.608750e+06	3608750.0	3608750.0	0.0	0.0	3.608749e+06	3.000000	3.608747e+06	3.608746e+06	3608744.0	360
mean	1.261245e+18	0.0	0.0	NaN	NaN	2.772129e+05	0.333333	2.771045e-07	1.000000e+00	0.0	
std	4.999962e+16	0.0	0.0	NaN	NaN	5.266132e+08	0.577350	5.264072e-04	7.444523e-04	0.0	
min	1.211050e+18	0.0	0.0	NaN	NaN	0.000000e+00	0.000000	0.000000e+00	0.000000e+00	0.0	
25%	1.211050e+18	0.0	0.0	NaN	NaN	0.000000e+00	0.000000	0.000000e+00	1.000000e+00	0.0	
50%	1.311050e+18	0.0	0.0	NaN	NaN	0.000000e+00	0.000000	0.000000e+00	1.000000e+00	0.0	
75%	1.311050e+18	0.0	0.0	NaN	NaN	0.000000e+00	0.500000	0.000000e+00	1.000000e+00	0.0	
max	1.311050e+18	0.0	0.0	NaN	NaN	1.000392e+12	1.000000	1.000000e+00	2.000000e+00	0.0	

Figure 3 Summary of statistics from various data, segmented for evidence purposes.

4.3. Data understanding:

In this phase of the CRISP-DM framework, the data sources studied, and the necessary data collected. The data labels, as identified in Tables 5 and 6, established to facilitate further analysis.

4.3.1. Study of Data Sources

In this phase, the researcher carefully examined the data sources, which include transactional data from mobile banking and mobile money services. The data sources evaluated for their relevance and completeness in relation to the objectives of the study. The researcher identified the structure, types of variables, and the overall quality of the data.

The mobile banking data, for instance, included fields such as transaction ID, timestamp, sender and recipient details, and transaction amounts. Similarly, the mobile money data had attributes that were essential for detecting fraudulent activities. This study of data sources is critical as it influences the subsequent steps in data preparation and modeling.

4.3.2. Identification of Data Labels

The identification of data labels is a fundamentals step in developing supervised learning models, particularly in the context of the SVM-based fraud detection model utilized in this study. Data labels serve as the ground truth that the model uses during training to learn how to classify transactions as either 'fraudulent' or 'legitimate.'

In this research, the data labels derived from comprehensive mobile banking and mobile money datasets provided by the bank. These datasets included transactions that had already been classified through historical analysis and prior investigations conducted by the bank's fraud management and audit teams. Transactions labelled based on whether they had previously identified as fraudulent or legitimate. This labeling was crucial because it provided the model with the necessary examples to learn from it.

4.4. Data preparation

4.4.1. Overview of data preparation

During the data preparation phase, the final dataset was constructed from the initial raw data collected from mobile banking and mobile money sources. This phase involved several critical tasks, including data cleaning, feature selection, and transforming the data into a format suitable for modelling. These activities were iterative and often revisited throughout the process, rather than following a strict sequential order (Rüdiger Wirth, 2017).

Data Fields and Initial Processing:

The mobile banking transaction dataset included a variety of fields, such as the transaction channel (e.g., USSD, Android, iPhone), currency type, transaction amounts, sources and beneficiaries, transaction dates, and statuses (e.g., cancelled, paid). Transactions were categorized into various types, such as Top-up, CBE-Birr, and others, including those associated with fraudulent activities on these digital platforms.

4.4.2. Data Pre-processing:

The data pre-processing phase was comprehensive, involving key steps such as importing CSV files from the mobile banking database, extracting relevant details from various fields, converting data types, filtering specific rows and columns, and restructuring tables for subsequent analysis. For instance, the SELECTED_OPTIONS column contained critical information such as the recipient's phone number and the transaction amount, which extracted using Python regular expressions, as detailed in Appendix B.

4.4.3. Feature Selection and Data Cleaning

During data preparation, some fields containing redundant or nearly identical information removed to streamline the analysis. These fields included CHANNEL_NAME, PRODUCT_ID, PROCESSED_DATE and SELECTED_OPTIONS in mobile banking dataset (relevant data set extracted from these fields like transaction ID, transaction types etc.) other data are removed as they are not appropriate in this study, which some of the fields are having zero values as

which in the screen shoot presented above. Whereas the in the cases of the mobile money TRANS_STATUS, TRANS_INITATE_TIME, TRANS_END_TIME, ORG_AMOUNT, ACTUAL_AMOUNT, BANK_ACCOUNT_NUMBER, BANK_ACCOUNT_NAME are used. The PRODUCT_NAME column provided crucial details on transaction types, such as rent payments, debt settlements, service fees, gifts, personal transfers, and mobile top-ups, along with additional data fields like amount and reference in mobile banking data set.

To determine which feature (column) from the dataset is next valuable procedure, it is mandatory to assess the usefulness of each column in terms of its potential impact on analysis or predictive modelling. Here is a brief evaluation of some key columns in addition to the once given in the data understanding section, which works for both the mobile banking and mobile money data set:

1. **AMOUNT extracted from mobile banking data set and REQUEST_AMOUNT in the mobile money data set (columns labelled amount this his study)**: This column represents the monetary value of the transaction. It is crucial for understanding the scale and impact of transactions. For predictive modeling, AMOUNT is often a strong feature as it directly relates to the target variable in financial analyses, such as predicting fraud or customer behavior.
2. **For mobile banking CUSTOMER_ID extracted from the column labelled PRODUCT_ID and for mobile money DEBIT_PARTY_ID, DEBIT_PARTY_ACCOUNT, CREDIT_PARTY_ID, CREDIT_PARTY_ACCOUNT**: This columns identifies the customers involved in the transaction as a credited party. However, it less useful for predictive models if used directly without aggregation because it can be too granular and lead to overfitting hence rejected.
3. **PROCESSED_DATE and TRANS_END_TIME** : This column provides the timestamp when the transaction was processed in both cases where they are duplicated in other columns. It is valuable for time-series analysis, understanding transaction patterns over time, and detecting anomalies or trends.
4. **TRANSACTION_TYPE**: This column indicates the nature of the transaction (e.g., debit, credit). It helps in distinguishing between different types of transactions,

which is useful for fraud detection and understanding transaction behaviour but in the data, the column holds same data as debit.

5. **STATUS_CODE and TRANS_STATUS:** This indicates the status of the transaction (e.g., paid, cancelled) same true for the mobile banking data set. It is important for tracking the outcome of transactions and can be useful for analysing transaction success rates and detecting issues but not used for feature in this study and already cleaned from the data under study in the cleaning phase.
6. **Transaction ID for mobile banking data set and ORDERID for mobile money data set**: this helps identify every transaction value and is only helps for investigation after the transactions are predicted and the data value is new for not every transaction hence can be used for feature selection in any way.

4.4.4. Feature for Predictive Modeling:

Among these features, **AMOUNT** is the most valuable for predictive modeling, particularly in fraud detection and financial analysis this is because:

- **Quantitative Relevance**: AMOUNT directly affects the analysis of transaction behaviour and financial modelling from the data set in the study . It is a continuous variable that can provide insights into transaction patterns, spending behavior, and anomalies.
- **Predictive Power**: In fraud detection, the amount of money involved in a transaction can be a strong indicator of fraudulent activity. Large or unusual transactions often warrant further investigation in this case.
- **Correlation with Target Variable**: AMOUNT is likely to have a significant correlation with the target variables (e.g., fraud or transaction type) in predictive models. For instance, unusually large transactions flagged as potential fraud.

To visualize the amount as a feature using a box plot for the transaction amounts based on the target (fraudulent vs. non-fraudulent transactions) is to visually and statistically assess the differences in transaction distributions between the two classes amount and a target variable which is created with the pattern distributed similar using the script below.

```
import matplotlib.pyplot as plt
import seaborn as sns

# Assuming df1 contains the data with 'Amount' and 'Target' columns
plt.figure(figsize=(10, 6))
sns.boxplot(x='Target', y='Amount', data=df1, palette='coolwarm')
plt.title('Distribution of Transaction Amounts by Target')
plt.xlabel('Target (0 = Non-Fraud, 1 = Fraud)')
plt.ylabel('Transaction Amount')
plt.show()
```

Figure 4 Script for the box plot of as amount is a variable

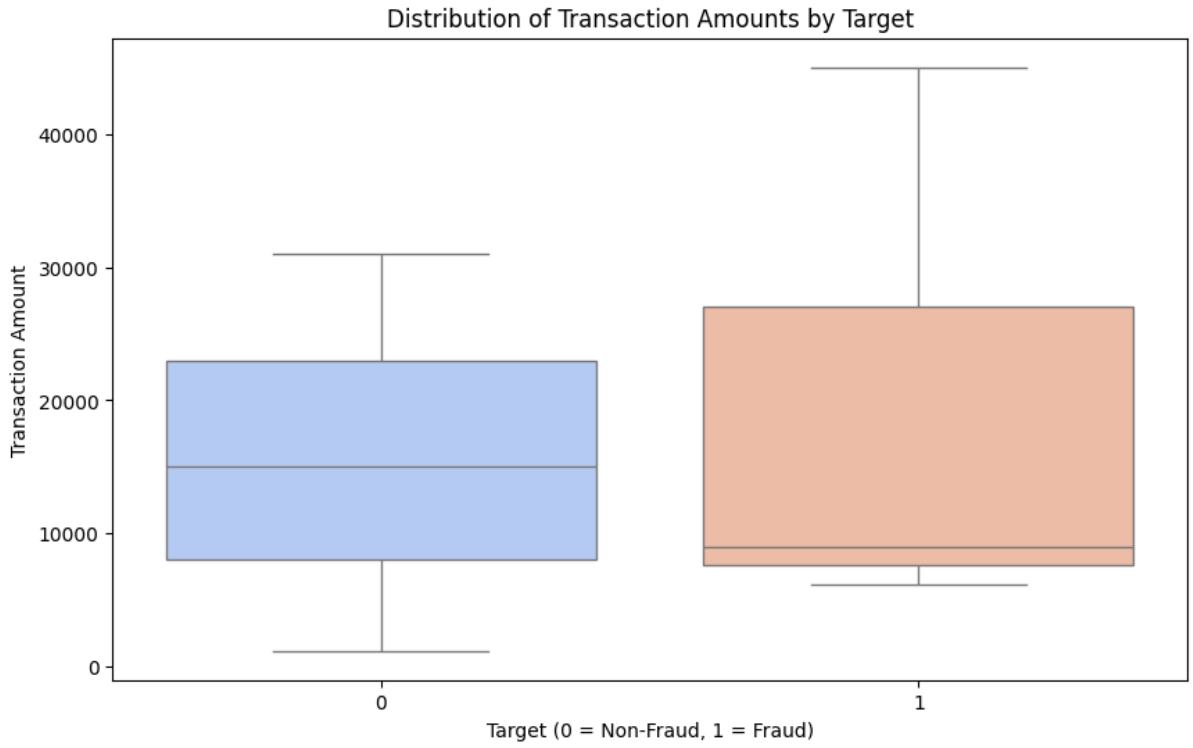


Figure 5 Box Plot of the transaction amount by the target

Explanation of the Plot

- **Non-Fraudulent Transactions (Target = 0)**:
  - The median transaction amount is relatively high.
  - The distribution is somewhat symmetric, suggesting that most transaction amounts fall within a middle range.

- The spread is fairly large, indicating variability in transaction amounts for non-fraudulent transactions.

- **Fraudulent Transactions (Target = 1)**:
  - The median transaction amount is lower compared to non-fraudulent transactions.
  - The IQR is larger, showing greater variability in the transaction amounts among fraudulent transactions.
  - The presence of high-value transactions (higher whiskers) might indicate that some fraudulent activities involve large amounts of money, though the majority are lower than non-fraudulent ones.

The clear difference in the distribution of transaction amounts between fraudulent and non-fraudulent transactions suggests that the transaction amount is a strong candidate for a predictive feature. Given the variance in median and spread across the two classes, leveraging threshold-based rules or integrating transaction amount into a machine-learning model could aid in distinguishing between fraudulent and non-fraudulent transactions. Therefore, in this study, the transaction amount is utilized as a feature within the context of the research.

4.4.5. Feature Engineering Considerations:

To analyse the feature using the data set at hand in this study the focus was on two main aspects: binning and interaction terms. Below is breakdown of how implemented:

Binning

Binning is the process of converting continuous variables into categorical ones by grouping the continuous values into "bins." This is for reducing noise and capturing the relationship between the feature and the target variable in a more interpretable manner. The bins constructed from

- **Amount:**
  - Create bins for transaction amounts, such as:
    - Low: 0-5000
    - Medium: 5001-20,000
    - High: 20001-50,000

- Very High: 50,001+

This can help in analysing how different ranges of transaction amounts influence other variables like STATUS_CODE or MOBILE_APP_ID but this study focus only on the processed date field data set to see time series analysis against the amount as a target for fraud detection.

The result of this process is a more structured dataset with features that are easier to interpret and potentially more informative for the model. This allows the SVM model to learn patterns based on categorized transaction amounts, which can be more effective than learning directly from raw transaction amounts.

In the code snippet below, the Amount values do not match exactly with the bin edges. This is because the Amount values are continuous data points, and the bins are defined ranges to categorize these continuous values.

Binning reduces the complexity of continuous data (amount) by aggregating it into meaningful categories. This can make it easier to analyse and understand patterns.

Reduces Noise: Continuous data can be noisy and include small variations that might not be significant. Binning smooths out these variations and highlights broader trends. Using the python code

- **PROCESSED_DATE:**
    - Extract time-based features like:
        - Hour of the day (morning, afternoon, evening, night)

- Day of the week (weekday vs. weekend)

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Sample DataFrame
data = {
    'MOBILE_APP_ID': [1, 2, 3, 4, 5, 6],
    'Amount': [1200, 6200, 31000, 9000, 15000, 45000],
    'Target': [0, 1, 0, 1, 0, 1]# The pattern [0, 1, 0, 1, 0, 1] provides an
    #equal number of samples for each class, which is useful for ensuring that the model is tested on a balanced dataset.
}



# Define bins and categorize the amounts
bins = [0, 5000, 20000, 50000, float('inf')]
labels = ['Low', 'Medium', 'High', 'Very High']

df['Amount_Binned'] = pd.cut(df['Amount'], bins=bins, labels=labels, right=False)

# Count the number of transactions in each bin
bin_counts = df['Amount_Binned'].value_counts().sort_index()

# Plot the distribution of transactions across bins
plt.figure(figsize=(8, 6))
sns.barplot(x=bin_counts.index, y=bin_counts.values, palette='viridis')
plt.xlabel('Transaction Amount Bin')
plt.ylabel('Number of Transactions')
plt.title('Distribution of Transactions Across Amount Bins')
plt.show()
```

Figure 6 Binning to convert continues amount data into category for SVM training

- Binning time-based features can help understand patterns in transaction times, such as p

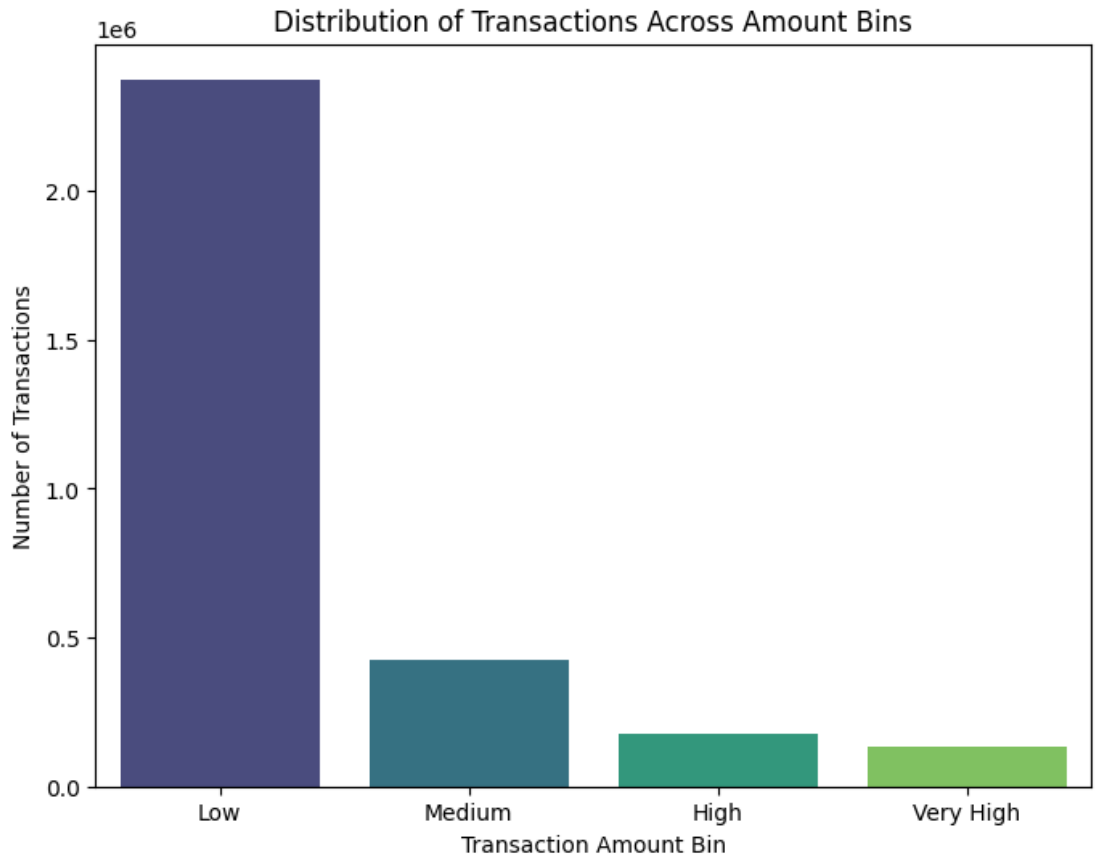


Figure 7 how the transaction amounts are categorized and distributed among the bins.

The main purpose of this code is to provide a comprehensive visualization of the daily transaction amounts and the distribution of transactions across different amount bins over time. This type of analysis can be useful for understanding the patterns and trends in the transaction data, which could potentially inform business decisions or help identify any anomalies or changes in the transaction behavior.

```python
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.dates import DateFormatter, WeekdayLocator, DayLocator, AutoDateLocator

# Convert 'PROCESSED_DATE' to datetime
df['PROCESSED_DATE'] = pd.to_datetime(df['PROCESSED_DATE'])

# Set 'PROCESSED_DATE' as index
df = df.set_index('PROCESSED_DATE')

# Resample data to daily frequency and sum the amounts
daily_amount = df['Amount'].resample('D').sum()

# Resample and count occurrences of each bin
daily_binned_counts = df.groupby([pd.Grouper(freq='D'), 'Amount_Binned']).size().unstack(fill_value=0)

# Plot time series
fig, ax1 = plt.subplots(figsize=(12, 6))

# Plot total daily amount
ax1.plot(daily_amount.index, daily_amount.values, color='blue', label='Total Amount')

# Set x-axis tick format
date_formatter = DateFormatter('%y-%m-%d %H:%M:%S')
ax1.xaxis.set_major_formatter(date_formatter)
plt.xticks(rotation=45)

# Create a second y-axis for binned counts
ax2 = ax1.twinx()
daily_binned_counts.plot(kind='bar', ax=ax2, alpha=0.5, width=0.7)
ax2.set_ylabel('Number of Transactions in Each Bin')
ax2.legend(loc='upper left')

# Adjust x-axis to show weekly ticks
locator = AutoDateLocator()
ax1.xaxis.set_major_locator(locator)

plt.title('Time Series Analysis of Transaction Amounts and Bins')
plt.tight_layout()
plt.show()
```

The provided code performs a time series analysis of transaction amounts and their distribution across different transaction amount bins. Below is a breakdown of what the code does:

1. Data Preparation:

- The code converts the 'PROCESSED_DATE' column to date time format.

- It sets the 'PROCESSED_DATE' column as the index of the Data Frame.

2. Daily Transaction Amount Summary:

- The code resamples the data to a daily frequency and sums the 'Amount' column, resulting in a series called 'daily_amount' that represents the total daily transaction amount.

3. Daily Binned Transactions Count:

- The code groups the data by the daily frequency (using 'pd.Grouper(freq='D')') and the 'Amount_Binned' column.

- It then counts the number of transactions in each bin for each day, creating a DataFrame called 'daily_binned_counts'.

4. Visualization:

- The code creates a figure with two y-axes:

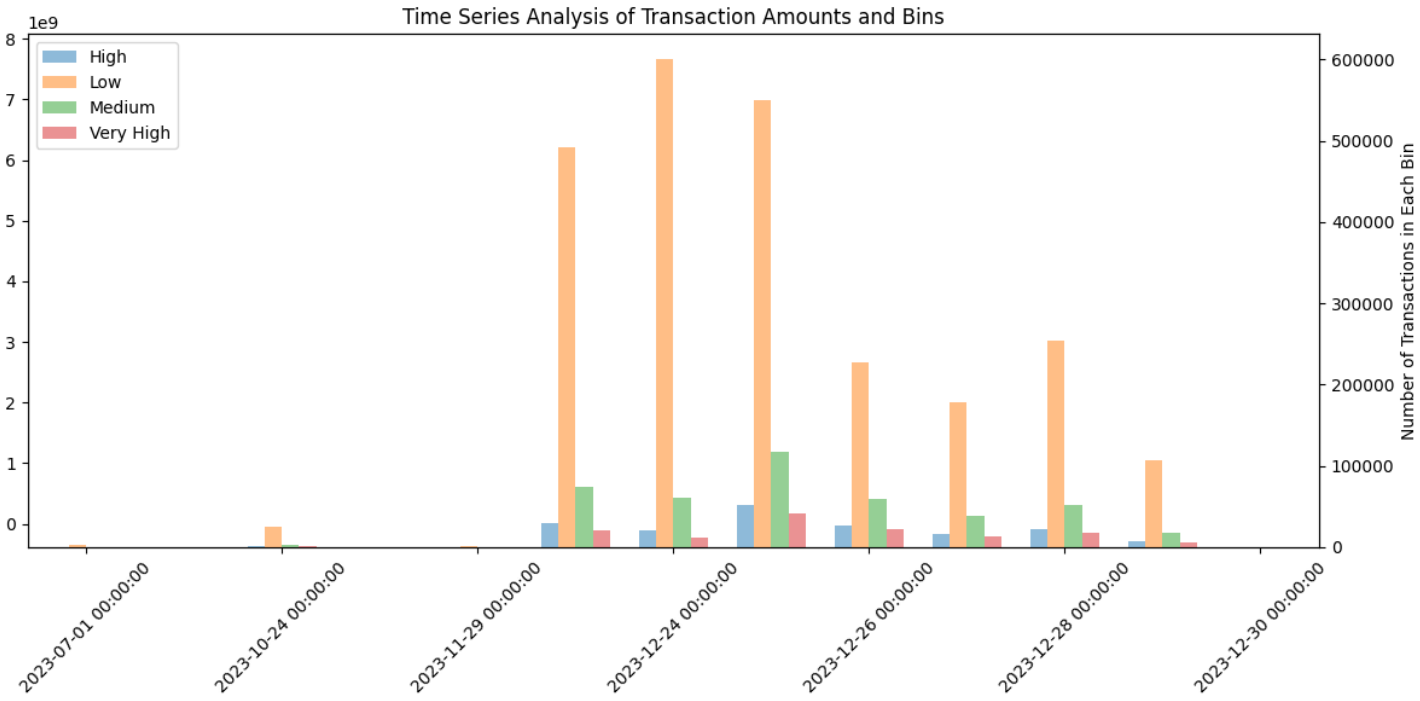
- The left y-axis plots the 'daily_amount' time series in blue.

- The right y-axis plots the 'daily_binned_counts' as a stacked bar chart, showing the number of transactions in each bin over time.

- The x-axis tick format is set to display the date in the 'yy-mm-dd' format and the time in the 'HH:MM:SS' format.

- The x-axis ticks are rotated by 45 degrees for better readability.

- The plot is titled "Time Series Analysis of Transaction Amounts and Bins".



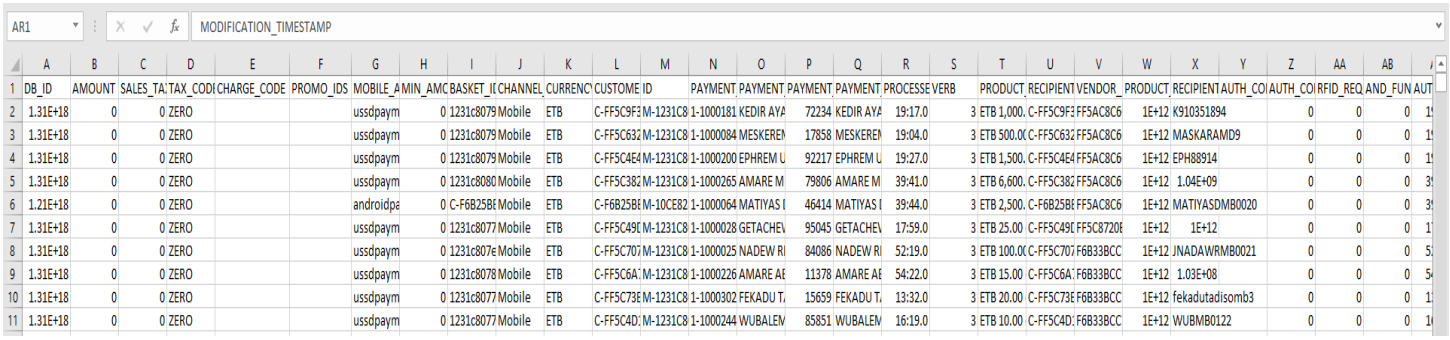
4.5. Data cleaning

The breakdown of the data and explanations for the removed transactions:

- **Missing or Incomplete Data:** The data cleaning process removed transactions that lacked essential information, such as transaction amounts, recipient details, or

timestamps. Inconsistent formatting (e.g., dates, transaction ID,) have been a reason for removal.

- **Duplicate Transactions:** The cleaning process have identified and removed duplicate entries for the same transaction. As the data owners explained because of the connection errors duplicate could be register in the database. This has been identified form the same transaction id.
- **Cancelled Transactions:** cancelled transactions have been excluded since they were not relevant for fraud detection or other analysis purposes in this particular case.
- **Invalid or Erroneous Data:** Transactions with invalid characters, nonsensical data points, or illogical values (e.g., negative transaction amounts but cancel transactions) have been removed during cleaning but the date owners say they don't still figure out way some data has a negative value. Others explain that some transaction has been presented on as negative since some of the transaction are not run instantly but in a batch mode.

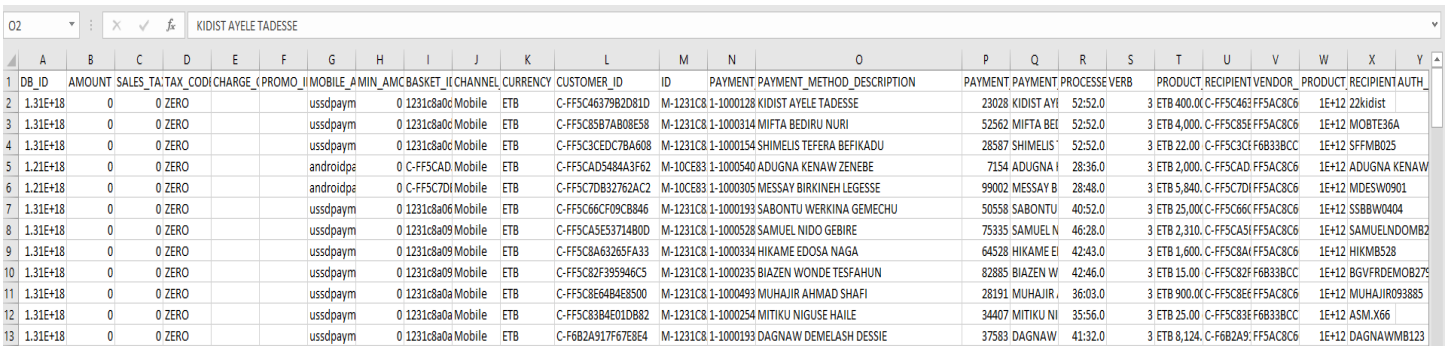


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	DB_ID	AMOUNT	SALES_TA	TAX_COD	CHARGE_CODE	PROMO_IDS	MOBILE_A	MIN_AMC	BASKET_I	CHANNEL	CURRENC	CUSTOME	ID	PAYMENT	PAYMENT	PAYMENT	PAYMENT	PROCESSE	VERB	PRODUCT	RECIPIENT	VENDOR_	PRODUCT	RECIPIENT	AUTH_CO	AUTH_CO	RFID_REQ	AND_FUN	AUT
2	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8079	Mobile	ETB	C-FF5C9F3	M-1231C8	1-1000181	KEDIR AYA	72234	KEDIR AYA	19:17.0	3	ETB 1,000.	C-FF5C9F3	FF5AC8C6	1E+12	K910351894		0	0	0	1
3	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8079	Mobile	ETB	C-FF5C632	M-1231C8	1-1000084	MESKEREM	17858	MESKEREM	19:04.0	3	ETB 500.0(	C-FF5C632	FF5AC8C6	1E+12	MASKARAMD9		0	0	0	1
4	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8079	Mobile	ETB	C-FF5C4E4	M-1231C8	1-1000200	EPHREM U	92217	EPHREM U	19:27.0	3	ETB 1,500.	C-FF5C4E4	FF5AC8C6	1E+12	EPH88914		0	0	0	1
5	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8080	Mobile	ETB	C-FF5C382	M-1231C8	1-1000265	AMARE M	79806	AMARE M	39:41.0	3	ETB 6,600.	C-FF5C382	FF5AC8C6	1E+12	1.04E+09		0	0	0	3
6	1.21E+18	0	0	ZERO			androidpa	0	C-F6B25BE	Mobile	ETB	C-F6B25BE	M-10CE82	1-1000064	MATIYAS I	46414	MATIYAS I	39:44.0	3	ETB 2,500.	C-F6B25BE	FF5AC8C6	1E+12	MATIYASDMB0020		0	0	0	3
7	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8077	Mobile	ETB	C-FF5C49(	M-1231C8	1-1000028	GETACHEV	95045	GETACHEV	17:59.0	3	ETB 25.00	C-FF5C49(	FF5C8720E	1E+12	1E+12		0	0	0	1
8	1.31E+18	0	0	ZERO			ussdpaym	0	1231c807e	Mobile	ETB	C-FF5C707	M-1231C8	1-1000025	NADEW RI	84086	NADEW RI	52:19.0	3	ETB 100.0(	C-FF5C707	F6B33BCC	1E+12	JNADAWRMB0021		0	0	0	5
9	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8078	Mobile	ETB	C-FF5C6A(	M-1231C8	1-1000226	AMARE AE	11378	AMARE AE	54:22.0	3	ETB 15.00	C-FF5C6A(	F6B33BCC	1E+12	1.03E+08		0	0	0	5
10	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8077	Mobile	ETB	C-FF5C73E	M-1231C8	1-1000302	FEKADU T	15659	FEKADU T	13:32.0	3	ETB 20.00	C-FF5C73E	F6B33BCC	1E+12	fekadutadisomb3		0	0	0	1
11	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8077	Mobile	ETB	C-FF5C4D(	M-1231C8	1-1000244	WUBALEM	85851	WUBALEM	16:19.0	3	ETB 10.00	C-FF5C4D(	F6B33BCC	1E+12	WUBMB0122		0	0	0	1

Figure 8 Jul012023MBTransaction Data

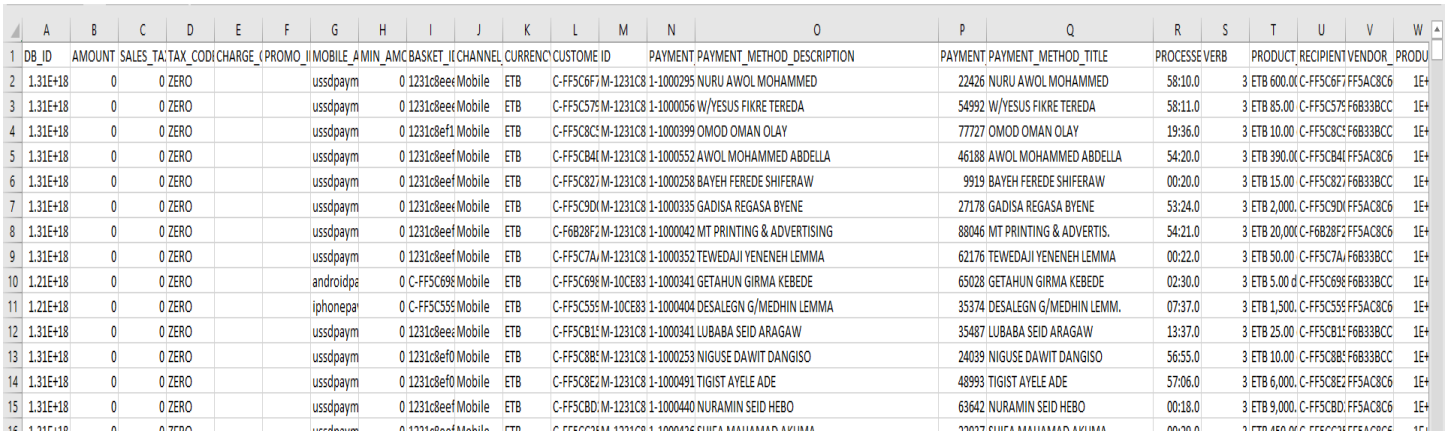
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	DB_ID	AMOUNT	SALES_TA	TAX_COD	CHARGE_(	PROMO_I	MOBILE_A	MIN_AMC	BASKET_ID	CHANNEL	CURRENC	CUSTOME	ID	PAYMENT	PAYMENT	PAYMENT	PAYMENT	PROCESSE	VERB	PRODUCT	RECIPIENT	VENDOR_	PRODUCT	RECIPIENT	AUTH_CO	AUTH_CO	RFID_REQ	AND_FU
2	1.31E+18	0	0	ZERO			ussdpaym	0	1231c85271921f36	Mobile	ETB	C-FF5C8BE	M-1231C8	1-1000431	TEMESGEN	92608	TEMESGEN	26:41.0	3	ETB 15,00(	C-FF5C8BE	FF5C5711E	1E+12	TEMMUU6996		0	0	
3	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8526fd490e3	Mobile	ETB	C-FF5C97E	M-1231C8	1-1000488	CHERINET	88352	CHERINET	21:49.0	3	ETB 15.00	C-FF5C97E	F6B33BCC	1E+12	1.02E+09		0	0	
4	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8526fd49129	Mobile	ETB	C-FF5C91E	M-1231C8	1-1000498	MUNIR M	98024	MUNIR M	21:49.0	3	ETB 25.00	C-FF5C91E	F6B33BCC	1E+12	1E+12		0	0	
5	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8526fd30a84	Mobile	ETB	C-F6B2FEE	M-1231C8	1-1000264	TUMEBO M	98627	TUMEBO M	21:46.0	3	Saving - E	C-F6B2FEE	FF5C5711E	1E+12	TT258MB001		0	0	
6	1.31E+18	0	0	ZERO			ussdpaym	0	1231c852710bd878	Mobile	ETB	C-FF5C64C	M-1231C8	1-1000344	SEUDI ESA	45047	SEUDI ESA	25:13.0	3	ETB 30.00	C-FF5C64C	F6B33BCC	1E+12	AA21259RPBT4		0	0	
7	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8526fd30a10	Mobile	ETB	C-FF5C4DI	M-1231C8	1-1000189	KASAHUN	13177	KASAHUN	21:48.0	3	ETB 25.00	C-FF5C4DI	F6B33BCC	1E+12	kumb123		0	0	
8	1.21E+18	0	0	ZERO			androidpa	0	C-FF5C588F92363E2-169	Mobile	ETB	C-FF5C588	M-10CE82	1-1000302	WEYNI HA	6093	WEYNI HA	24:21.0	3	ETB 1,950.	C-FF5C588	FF5AC8C6	1E+12	WMDMB1706		0	0	
9	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8520731c300	Mobile	ETB	C-FF5CB45	M-1231C8	1-1000062	ETALEMAH	39608	ETALEMAH	29:14.0	3	ETB 15.00	C-FF5CB45	F6B33BCC	1E+12	ETALEMAHU YETNEB		0	0	
10	1.21E+18	0	0	ZERO			androidpa	0	C-FF5CB7C0BB31A33-16	Mobile	ETB	C-FF5CB7(	M-10CE82	1-1000558	MERWAN	30767	MERWAN	24:20.0	3	Mudarab -	C-FF5CB7(	FF5C5711E	1E+12	MERWAN MEHAMED		0	0	
11	1.31E+18	0	0	ZERO			ussdpaym	0	1231c85271921ee0	Mobile	ETB	C-FF5C8A(	M-1231C8	1-1000326	ZULMEKAI	637	ZULMEKAI	26:41.0	3	ETB 5,150.	C-FF5C8A(	FF5AC8C6	1E+12	ZULAB100		0	0	
12	1.31E+18	0	0	ZERO			ussdpaym	0	1231c85271921eda	Mobile	ETB	C-FF5CAD	M-1231C8	1-1000534	ELIAS GETI	55389	ELIAS GETI	26:40.0	3	ETB 800.0(	C-FF5CAD	FF5AC8C6	1E+12	ELIAS GETERA EKO		0	0	
13	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8524e6462f4	Mobile	ETB	C-FF5C2F7	M-1231C8	1-1000096	AKLILU M	88833	AKLILU M	48:19.0	3	ETB 50,00(	C-FF5C2F7	FF5AC8C6	1E+12	AKM205		0	0	
14	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8524e62dbe1	Mobile	ETB	C-FF5C3EE	M-1231C8	1-1000274	KIDIST SEI	94077	KIDIST SEI	48:18.0	3	ETB 6,000.	C-FF5C3EE	FF5AC8C6	1E+12	K091201		0	0	
15	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8524e62dbe7	Mobile	ETB	C-FF5CB82	M-1231C8	1-1000558	DEREJE TC	14985	DEREJE TC	48:17.0	3	CHALTU U	C-FF5CB82	FF5AC8C6	1E+12	DEREJE TOLA DEBELA		0	0	
16	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8522c201c3e	Mobile	ETB	C-FF5C341	M-1231C8	1-1000150	A/SELAM /	93207	A/SELAM /	12:30.0	3	ETB 300.0(	C-FF5C341	FF5AC8C6	1E+12	ASELAM2180		0	0	
17	1.21E+18	0	0	ZERO			androidpa	0	C-FF5CBA833EF6143-169	Mobile	ETB	C-FF5CBA(	M-10CE82	1-1000563	KAMIL ME	9797	KAMIL ME	46:28.0	3	Wadiah -	C-FF5CBA(	FF5C3B63(	10005631(	KAMIL MEHAMMED I		0	0	

Figure 9 Aug.072023MBTransaction Data



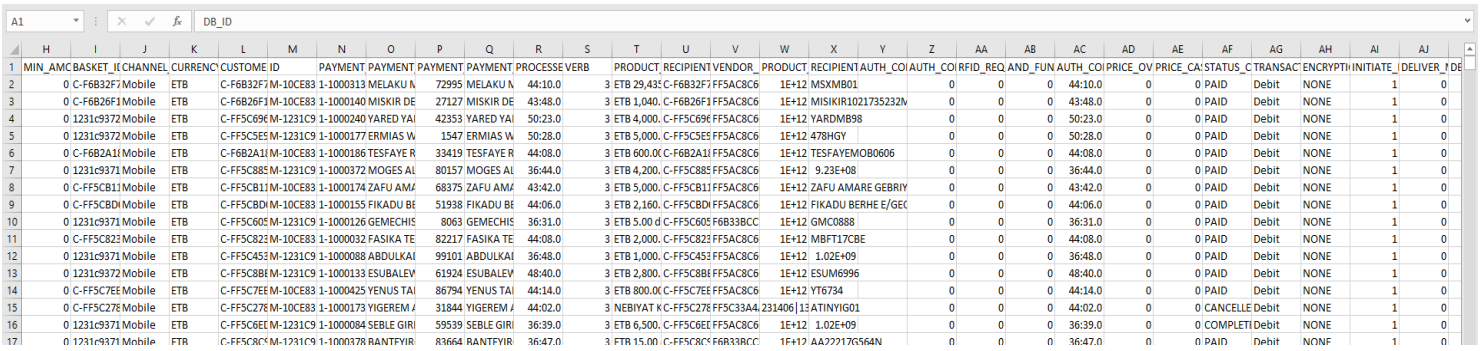
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	DB_ID	AMOUNT	SALES_TA	TAX_COD	CHARGE_(	PROMO_I	MOBILE_A	MIN_AMC	BASKET_I(	CHANNEL	CURRENCY	CUSTOMER_ID	ID	PAYMENT	PAYMENT_METHOD_DESCRIPTION	PAYMENT	PAYMENT	PROCESSE	VERB	PRODUCT_	RECIPIENT	VENDOR_	PRODUCT	RECIPIENT	AUTH_
2	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a0c	Mobile	ETB	C-FF5C46379B2D81D	M-1231C8	1-1000128	KIDIST AYELE TADESSE	23028	KIDIST AY(	52:52.0	3	ETB 400.0(	C-FF5C463	FF5AC8C6	1E+12	22kidist	
3	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a0c	Mobile	ETB	C-FF5C85B7AB08E58	M-1231C8	1-1000314	MIFTA BEDIRU NURI	52562	MIFTA BE(	52:52.0	3	ETB 4,000.	C-FF5C85E	FF5AC8C6	1E+12	MOBTE36A	
4	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a0c	Mobile	ETB	C-FF5C3CEDC7BA608	M-1231C8	1-1000154	SHIMELIS TEFERA BEFIKADU	28587	SHIMELIS '	52:52.0	3	ETB 22.00	C-FF5C3CE	F6B33BCC	1E+12	SFFMB025	
5	1.21E+18	0	0	ZERO			androidpa	0	C-FF5CAD	Mobile	ETB	C-FF5CAD5484A3F62	M-10CE83	1-1000540	ADUGNA KENAW ZENEBE	7154	ADUGNA I	28:36.0	3	ETB 2,000.	C-FF5CAD	FF5AC8C6	1E+12	ADUGNA KENAW	
6	1.21E+18	0	0	ZERO			androidpa	0	C-FF5C7D	Mobile	ETB	C-FF5C7DB32762AC2	M-10CE83	1-1000305	MESSAY BIRKINEH LEGESSE	99002	MESSAY B	28:48.0	3	ETB 5,840.	C-FF5C7D(	FF5AC8C6	1E+12	MDESW0901	
7	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a06	Mobile	ETB	C-FF5C66CF09CB846	M-1231C8	1-1000193	SABONTU WERKINA GEMECHU	50558	SABONTU	40:52.0	3	ETB 25,00(	C-FF5C66(	FF5AC8C6	1E+12	SSBBW0404	
8	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a09	Mobile	ETB	C-FF5CA5E53714B0D	M-1231C8	1-1000528	SAMUEL NIDO GEBIRE	75335	SAMUEL N	46:28.0	3	ETB 2,310.	C-FF5CA5(	FF5AC8C6	1E+12	SAMUELNDOMB2	
9	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a09	Mobile	ETB	C-FF5C8A63265FA33	M-1231C8	1-1000334	HIKAME EDOSA NAGA	64528	HIKAME EI	42:43.0	3	ETB 1,600.	C-FF5C8A(	FF5AC8C6	1E+12	HIKMB528	
10	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a09	Mobile	ETB	C-FF5C82F395946C5	M-1231C8	1-1000235	BIAZEN WONDE TESFAHUN	82885	BIAZEN W	42:46.0	3	ETB 15.00	C-FF5C82F	F6B33BCC	1E+12	BGVFRDEMOB275	
11	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a0a	Mobile	ETB	C-FF5C8E64B4E8500	M-1231C8	1-1000493	MUHAJIR AHMAD SHAFI	28191	MUHAJIR /	36:03.0	3	ETB 900.0(	C-FF5C8EE	FF5AC8C6	1E+12	MUHAJIR093885	
12	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a0a	Mobile	ETB	C-FF5C83B4E01D882	M-1231C8	1-1000254	MITIKU NIGUSE HAILE	34407	MITIKU NI	35:56.0	3	ETB 25.00	C-FF5C83E	F6B33BCC	1E+12	ASM.X66	
13	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8a0a	Mobile	ETB	C-F6B2A917F67E8E4	M-1231C8	1-1000193	DAGNAW DEMELASH DESSIE	37583	DAGNAW	41:32.0	3	ETB 8,124.	C-F6B2A9(	FF5AC8C6	1E+12	DAGNAWMB123	

Figure 11 Sept.152023MBTransaction Data



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	DB_ID	AMOUNT	SALES_TA	TAX_COD	CHARGE_(	PROMO_I	MOBILE_A	MIN_AMC	BASKET_I(	CHANNEL	CURRENC	CUSTOME	ID	PAYMENT	PAYMENT_METHOD_DESCRIPTION	PAYMENT	PAYMENT_METHOD_TITLE	PROCESSE	VERB	PRODUCT_	RECIPIENT	VENDOR_	PRODU
2	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8ee(	Mobile	ETB	C-FF5C6F7	M-1231C8	1-1000295	NURU AWOL MOHAMMED	22426	NURU AWOL MOHAMMED	58:10.0	3	ETB 600.0(	C-FF5C6F7	FF5AC8C6	1E+
3	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8ee(	Mobile	ETB	C-FF5C575	M-1231C8	1-1000056	W/YESUS FIKRE TEREDA	54992	W/YESUS FIKRE TEREDA	58:11.0	3	ETB 85.00	C-FF5C575	F6B33BCC	1E+
4	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8ef1	Mobile	ETB	C-FF5C8C5	M-1231C8	1-1000399	OMOD OMAN OLAY	77727	OMOD OMAN OLAY	19:36.0	3	ETB 10.00	C-FF5C8C5	F6B33BCC	1E+
5	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8eef	Mobile	ETB	C-FF5CB4(	M-1231C8	1-1000552	AWOL MOHAMMED ABDELLA	46188	AWOL MOHAMMED ABDELLA	54:20.0	3	ETB 390.0(	C-FF5CB4(	FF5AC8C6	1E+
6	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8eef	Mobile	ETB	C-FF5C827	M-1231C8	1-1000258	BAYEH FERede SHIFERAW	9919	BAYEH FERede SHIFERAW	00:20.0	3	ETB 15.00	C-FF5C827	F6B33BCC	1E+
7	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8eef	Mobile	ETB	C-FF5C9D(	M-1231C8	1-1000335	GADISA REGASA BYENE	27178	GADISA REGASA BYENE	53:24.0	3	ETB 2,000.	C-FF5C9D(	FF5AC8C6	1E+
8	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8eef	Mobile	ETB	C-F6B28F2	M-1231C8	1-1000042	MT PRINTING & ADVERTISING	88046	MT PRINTING & ADVERTIS.	54:21.0	3	ETB 20,00(	C-F6B28F2	FF5AC8C6	1E+
9	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8eef	Mobile	ETB	C-FF5C7A/	M-1231C8	1-1000352	TEWEDAJI YENENEH LEMMA	62176	TEWEDAJI YENENEH LEMMA	00:22.0	3	ETB 50.00	C-FF5C7A/	F6B33BCC	1E+
10	1.21E+18	0	0	ZERO			androidpa	0	C-FF5C698	Mobile	ETB	C-FF5C698	M-10CE83	1-1000341	GETAHUN GIRMA KEBEDE	65028	GETAHUN GIRMA KEBEDE	02:30.0	3	ETB 5.00 d	C-FF5C698	F6B33BCC	1E+
11	1.21E+18	0	0	ZERO			iphonepa	0	C-FF5C555	Mobile	ETB	C-FF5C555	M-10CE83	1-1000404	DESALEGN G/MEDHIN LEMMA	35374	DESALEGN G/MEDHIN LEMM.	07:37.0	3	ETB 1,500.	C-FF5C555	FF5AC8C6	1E+
12	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8eec	Mobile	ETB	C-FF5CB1E	M-1231C8	1-1000341	LUBABA SEID ARAGAW	35487	LUBABA SEID ARAGAW	13:37.0	3	ETB 25.00	C-FF5CB1E	F6B33BCC	1E+
13	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8ef0	Mobile	ETB	C-FF5C8B5	M-1231C8	1-1000253	NIGUSE DAWIT DANGISO	24039	NIGUSE DAWIT DANGISO	56:55.0	3	ETB 10.00	C-FF5C8B5	F6B33BCC	1E+
14	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8ef0	Mobile	ETB	C-FF5C8E2	M-1231C8	1-1000491	TIGIST AYELE ADE	48993	TIGIST AYELE ADE	57:06.0	3	ETB 6,000.	C-FF5C8E2	FF5AC8C6	1E+
15	1.31E+18	0	0	ZERO			ussdpaym	0	1231c8eef	Mobile	ETB	C-FF5CBD(	M-1231C8	1-1000440	NURAMIN SEID HEBO	63642	NURAMIN SEID HEBO	00:18.0	3	ETB 9,000.	C-FF5CBD(	FF5AC8C6	1E+

Figure 10 Oct.242023MBTransaction Data

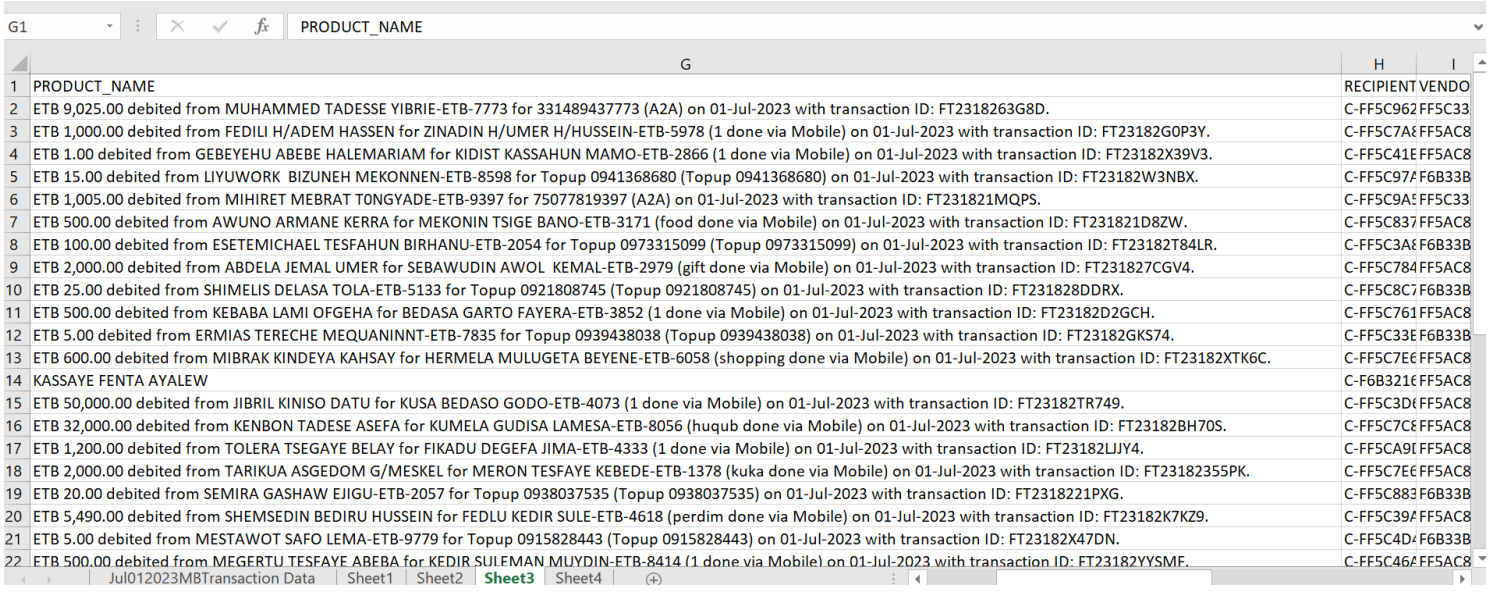


MIN_AMC	BASKET_IE	CHANNEL	CURRENC	CUSTOME	ID	PAYMENT	PAYMENT	PAYMENT	PAYMENT	PROCESSE	VERB	PRODUCT	RECIPIENT	VENDOR_	PRODUCT	RECIPIENT	AUTH_CO	AUTH_CO	RFID_REQ	AND_FUN	AUTH_CO	PRICE_OV	PRICE_CA	STATUS_C	TRANSAC	ENCRYPTI	INITIATE_	DELIVER_	DE
0	C-F6B32F7	Mobile	ETB	C-F6B32F7	M-10CE83	1-1000313	MELAKU N	72995	MELAKU N	44:10.0	3	ETB 29,435	C-F6B32F7	FF5AC8C6	1E+12	MSXMB01	0	0	0	0	44:10.0	0	0	PAID	Debit	NONE	1	0	
0	C-F6B26F1	Mobile	ETB	C-F6B26F1	M-10CE83	1-1000140	MISKIR DE	27127	MISKIR DE	43:48.0	3	ETB 1,040.	C-F6B26F1	FF5AC8C6	1E+12	MISIKIR1021735232M	0	0	0	0	43:48.0	0	0	PAID	Debit	NONE	1	0	
0	1231c9372	Mobile	ETB	C-FF5C69E	M-1231C9	1-1000240	YARED YAI	42353	YARED YAI	50:23.0	3	ETB 4,000.	C-FF5C69E	FF5AC8C6	1E+12	YARDMB98	0	0	0	0	50:23.0	0	0	PAID	Debit	NONE	1	0	
0	1231c9372	Mobile	ETB	C-FF5C5E5	M-1231C9	1-1000177	ERMIAS W	1547	ERMIAS W	50:28.0	3	ETB 5,000.	C-FF5C5E5	FF5AC8C6	1E+12	478HGY	0	0	0	0	50:28.0	0	0	PAID	Debit	NONE	1	0	
0	C-F6B2A1E	Mobile	ETB	C-F6B2A1E	M-10CE83	1-1000186	TESFAYE R	33419	TESFAYE R	44:08.0	3	ETB 600.0(	C-F6B2A1E	FF5AC8C6	1E+12	TESFAYEMOB0606	0	0	0	0	44:08.0	0	0	PAID	Debit	NONE	1	0	
0	1231c9371	Mobile	ETB	C-FF5C885	M-1231C9	1-1000372	MOGES AL	80157	MOGES AL	36:44.0	3	ETB 4,200.	C-FF5C885	FF5AC8C6	1E+12	9.23E+08	0	0	0	0	36:44.0	0	0	PAID	Debit	NONE	1	0	
0	C-FF5CB11	Mobile	ETB	C-FF5CB11	M-10CE83	1-1000174	ZAFU AMA	68375	ZAFU AMA	43:42.0	3	ETB 5,000.	C-FF5CB11	FF5AC8C6	1E+12	ZAFU AMARE GEBRIY	0	0	0	0	43:42.0	0	0	PAID	Debit	NONE	1	0	
0	C-FF5CBDI	Mobile	ETB	C-FF5CBDI	M-10CE83	1-1000155	FIKADU BE	51938	FIKADU BE	44:06.0	3	ETB 2,160.	C-FF5CBDI	FF5AC8C6	1E+12	FIKADU BERHE E/GEC	0	0	0	0	44:06.0	0	0	PAID	Debit	NONE	1	0	
0	1231c9371	Mobile	ETB	C-FF5C605	M-1231C9	1-1000126	GEMECHIS	8063	GEMECHIS	36:31.0	3	ETB 5.00 d	C-FF5C605	F6B33BCC	1E+12	GMC0888	0	0	0	0	36:31.0	0	0	PAID	Debit	NONE	1	0	
0	C-FF5C823	Mobile	ETB	C-FF5C823	M-10CE83	1-1000032	FASIKA TE	82217	FASIKA TE	44:08.0	3	ETB 2,000.	C-FF5C823	FF5AC8C6	1E+12	MBFT17CBE	0	0	0	0	44:08.0	0	0	PAID	Debit	NONE	1	0	
0	1231c9371	Mobile	ETB	C-FF5C453	M-1231C9	1-1000088	ABDULKAI	99101	ABDULKAI	36:48.0	3	ETB 1,000.	C-FF5C453	FF5AC8C6	1E+12	1.02E+09	0	0	0	0	36:48.0	0	0	PAID	Debit	NONE	1	0	
0	1231c9372	Mobile	ETB	C-FF5C8BE	M-1231C9	1-1000133	ESUBALEV	61924	ESUBALEV	48:40.0	3	ETB 2,800.	C-FF5C8BE	FF5AC8C6	1E+12	ESUM6996	0	0	0	0	48:40.0	0	0	PAID	Debit	NONE	1	0	
0	C-FF5C7EE	Mobile	ETB	C-FF5C7EE	M-10CE83	1-1000425	YENUS TAI	86794	YENUS TAI	44:14.0	3	ETB 800.0(	C-FF5C7EE	FF5AC8C6	1E+12	YT6734	0	0	0	0	44:14.0	0	0	PAID	Debit	NONE	1	0	
0	C-FF5C278	Mobile	ETB	C-FF5C278	M-10CE83	1-1000173	YIGEREM A	31844	YIGEREM A	44:02.0	3	NEBIYAT K	C-FF5C278	FF5C33A4	231406	13	ATINYIG01	0	0	0	0	44:02.0	0	0	CANCELLE	Debit	NONE	1	0
0	1231c9371	Mobile	ETB	C-FF5C6EE	M-1231C9	1-1000084	SEBLE GIR	59539	SEBLE GIR	36:39.0	3	ETB 6,500.	C-FF5C6EE	FF5AC8C6	1E+12	1.02E+09	0	0	0	0	36:39.0	0	0	COMPLETI	Debit	NONE	1	0	
0	1231c9371	Mobile	ETB	C-FF5C8C9	M-1231C9	1-1000378	BANTEYIR	83664	BANTEYIR	36:47.0	3	ETB 15.00	C-FF5C8C9	F6B33BCC	1E+12	AA22217G564N	0	0	0	0	36:47.0	0	0	PAID	Debit	NONE	1	0	

Figure 12 Nov29.242023MBTransaction Data

The PRODUCT_NAME field appears to contain structured information about a transaction, which requires parsing to get transaction necessary details for data analysis, including:

- Transaction Amount: In Ethiopian Birr (ETB)
- Debited Account Holder: Name of the account holder
- Recipient: Name or account number of the recipient
- Transaction Reason: Describes the purpose (e.g., gift, top-up, shopping among a few more)
- Transaction Channel: Indicates how the transaction was initiated (e.g., via Mobile)
- Transaction ID: A unique identifier for the transaction



	G	H	I
1	PRODUCT_NAME	RECIPIENT	VENDO
2	ETB 9,025.00 debited from MUHAMMED TADESSE YIBRIE-ETB-7773 for 331489437773 (A2A) on 01-Jul-2023 with transaction ID: FT2318263G8D.	C-FF5C962	FF5C33
3	ETB 1,000.00 debited from FEDILI H/ADEM HASSEN for ZINADIN H/UMER H/HUSSEIN-ETB-5978 (1 done via Mobile) on 01-Jul-2023 with transaction ID: FT23182G0P3Y.	C-FF5C7A8	FF5AC8
4	ETB 1.00 debited from GEBEYEHU ABEBE HALEMARIAM for KIDIST KASSAHUN MAMO-ETB-2866 (1 done via Mobile) on 01-Jul-2023 with transaction ID: FT23182X39V3.	C-FF5C41E	FF5AC8
5	ETB 15.00 debited from LIYUWORK BIZUNEH MEKONNEN-ETB-8598 for Topup 0941368680 (Topup 0941368680) on 01-Jul-2023 with transaction ID: FT23182W3NBX.	C-FF5C97A	F6B33B
6	ETB 1,005.00 debited from MIHIRET MEBRAT TONGYADE-ETB-9397 for 75077819397 (A2A) on 01-Jul-2023 with transaction ID: FT231821MQPS.	C-FF5C9A5	FF5C33
7	ETB 500.00 debited from AWUNO ARMANE KERRA for MEKONIN TSIGE BANO-ETB-3171 (food done via Mobile) on 01-Jul-2023 with transaction ID: FT231821D8ZW.	C-FF5C837	FF5AC8
8	ETB 100.00 debited from ESETEMICHAEL TESFAHUN BIRHANU-ETB-2054 for Topup 0973315099 (Topup 0973315099) on 01-Jul-2023 with transaction ID: FT23182T84LR.	C-FF5C3A8	F6B33B
9	ETB 2,000.00 debited from ABDELA JEMAL UMER for SEBAWUDIN AWOL KEMAL-ETB-2979 (gift done via Mobile) on 01-Jul-2023 with transaction ID: FT231827CGV4.	C-FF5C784	FF5AC8
10	ETB 25.00 debited from SHIMELIS DELASA TOLA-ETB-5133 for Topup 0921808745 (Topup 0921808745) on 01-Jul-2023 with transaction ID: FT231828DDRX.	C-FF5C8C7	F6B33B
11	ETB 500.00 debited from KEBABA LAMI OFGEHA for BEDASA GARTO FAYERA-ETB-3852 (1 done via Mobile) on 01-Jul-2023 with transaction ID: FT23182D2GCH.	C-FF5C761	FF5AC8
12	ETB 5.00 debited from ERMIAS TERECHE MEQUANINNT-ETB-7835 for Topup 0939438038 (Topup 0939438038) on 01-Jul-2023 with transaction ID: FT23182GKS74.	C-FF5C33E	F6B33B
13	ETB 600.00 debited from MIBRAK KINDEYA KAHSAY for HERMELA MULUGETA BEYENE-ETB-6058 (shopping done via Mobile) on 01-Jul-2023 with transaction ID: FT23182XTK6C.	C-FF5C7E6	FF5AC8
14	KASSAYE FENTA AYALEW	C-F6B321E	FF5AC8
15	ETB 50,000.00 debited from JIBRIL KINISO DATU for KUSA BEDASO GODO-ETB-4073 (1 done via Mobile) on 01-Jul-2023 with transaction ID: FT23182TR749.	C-FF5C3D(	FF5AC8
16	ETB 32,000.00 debited from KENBON TADESE ASEFA for KUMELA GUDISA LAMESA-ETB-8056 (huqub done via Mobile) on 01-Jul-2023 with transaction ID: FT23182BH70S.	C-FF5C7C8	FF5AC8
17	ETB 1,200.00 debited from TOLERA TSEGAYE BELAY for FIKADU DEGEFA JIMA-ETB-4333 (1 done via Mobile) on 01-Jul-2023 with transaction ID: FT23182LJJY4.	C-FF5CA9E	FF5AC8
18	ETB 2,000.00 debited from TARIKUA ASGEDOM G/MESKEL for MERON TESFAYE KEBEDE-ETB-1378 (kuka done via Mobile) on 01-Jul-2023 with transaction ID: FT23182355PK.	C-FF5C7E6	FF5AC8
19	ETB 20.00 debited from SEMIRA GASHAW EJIGU-ETB-2057 for Topup 0938037535 (Topup 0938037535) on 01-Jul-2023 with transaction ID: FT2318221PXG.	C-FF5C883	F6B33B
20	ETB 5,490.00 debited from SHEMSEDIN BEDIRU HUSSEIN for FEDLU KEDIR SULE-ETB-4618 (perdim done via Mobile) on 01-Jul-2023 with transaction ID: FT23182K7KZ9.	C-FF5C39A	FF5AC8
21	ETB 5.00 debited from MESTAWOT SAFO LEMA-ETB-9779 for Topup 0915828443 (Topup 0915828443) on 01-Jul-2023 with transaction ID: FT23182X47DN.	C-FF5C4D4	F6B33B
22	ETB 500.00 debited from MEGERTU TESFAYE ABEBA for KEDIR SULEMAN MUYDIN-ETB-8414 (1 done via Mobile) on 01-Jul-2023 with transaction ID: FT23182YYSMF.	C-FF5C46A	FF5AC8

Figure 13 Mobile banking data before processing in the product name column

The code to extract the data in this column using the code

```
import pandas as pd
import re

# Data extraction from the following data sample as shown in the excel
data = {
    "PRODUCT_NAME": [
        "ETB 9,025.00 debited from MUHAMMED TADESSE YIBRIE-ETB-7773 for 331489437773 (A2A) on 01-Jul-2023 with transaction ID: FT2318263G8D.",
        "ETB 1,000.00 debited from FEDILI H/ADEM HASSEN for ZINADIN H/UMER H/HUSSEIN-ETB-5978 (1 done via Mobile) on 01-Jul-2023 with transaction ID: FT23182G0P3Y.",
        # ... other data
    ]
}

df = pd.DataFrame(data)

def extract_info(text):
  """Extracts information from the text using regular expressions."""

  pattern = r"ETB (\d+,\d+\.\d+) debited from (.+) for (.+) \((\w+)\) on (\d{2}-\w{3}-\d{4}) with transaction ID: (FT\d+)"
  match = re.match(pattern, text)

  if match:
    amount, debited_account, recipient, reason, date, transaction_id = match.groups()
    return pd.Series([amount, debited_account, recipient, reason, date, transaction_id])
  else:
    return None

# Apply extraction to the "PRODUCT_NAME" column
df[['amount', 'debited_account', 'recipient', 'reason', 'date', 'transaction_id']] = df['PRODUCT_NAME'].apply(extract_info).tolist()

print(df)
```

The result of the code execution sample for further analysis

A	B	C	D	E	F
Transaction ID	Amount (ETB)	Debited Account Holder	Recipient	Reason	Transaction Channel
FT2318263G8D	9025	MUHAMMED TADESSE YIBRIE	331489437773	A2A	Mobile
FT23182G0P3Y	1000	FEDILI H/ADEM HASSEN	ZINADIN H/UMER H/HUSSEIN	P2P	Mobile
FT23182X39V3	1	GEBEYEHU ABEBE HALEMARIAM	KIDIST KASSAHUN MAMO	P2P	Mobile
FT23182W3NBX	15	LIYUWORK BIZUNEH MEKONNEN	941368680	Top-up	Mobile
FT231821MQPS	1005	MIHIRET MEBRAT T0NGYADE	75077819397	A2A	Mobile
FT231821D8ZW	500	AWUNO ARMANE KERRA	MEKONIN TSIGE BANO	Food	Mobile
FT23182T84LR	100	ESETEMICHAEL TESFAHUN BIRHANU	973315099	Top-up	Mobile
FT231827CGV4	2000	ABDELA JEMAL UMER	SEBAWUDIN AWOL KEMAL	Gift	Mobile
FT231828DDRX	25	SHIMELIS DELASA TOLA	921808745	Top-up	Mobile
FT23182D2GCH	500	KEBABA LAMI OFGEHA	BEDASA GARTO FAYERA	P2P	Mobile

This Graph shown below (Fig 4.) the number of mobile banking transactions captured before and after data cleaning for a period from July 1, 2023, to December 31, 2023. The data cleaning process removed a significant portion of transactions that result in a total reduction of 6,173,497 transactions (21.97%).

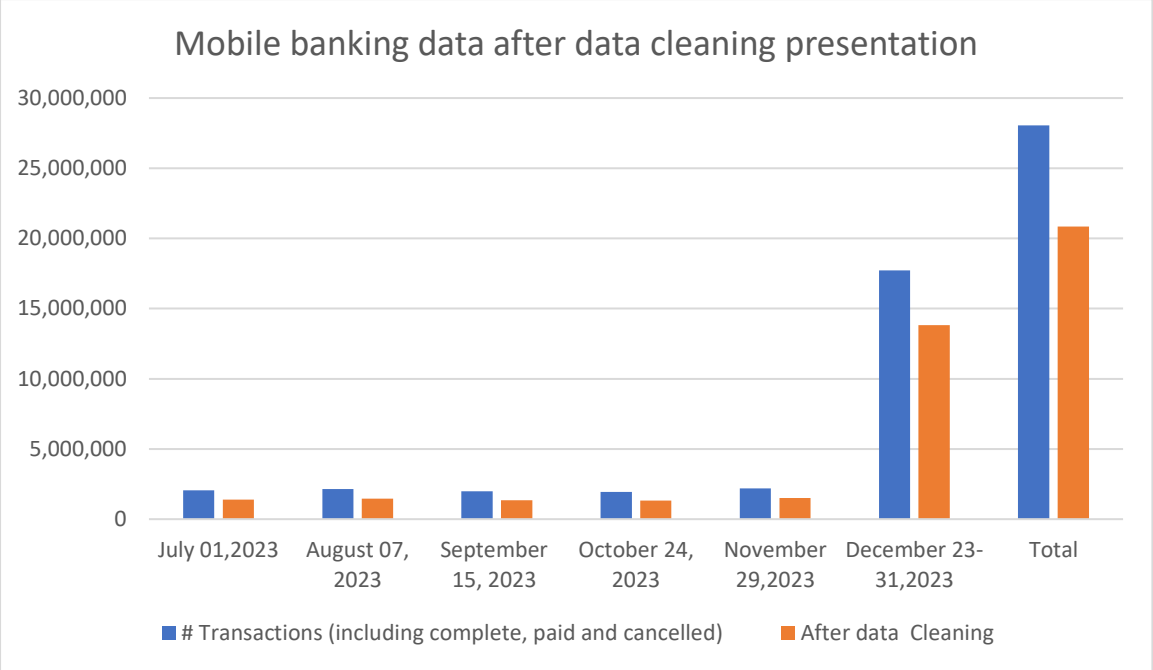


Figure 14 Mobile banking transactional data before and after date cleaning

4.5.1. Mobile money

Among the contents of this 46 dataset are varied details concerning different cellular money transactions, each column signifying a certain angle towards the transaction: giving a complete view of financial activity. Upon analysis of the data set of the following columns has been revoked for this analysis as they are not used as test cases or indirectly associated for this fraud detection agenda.

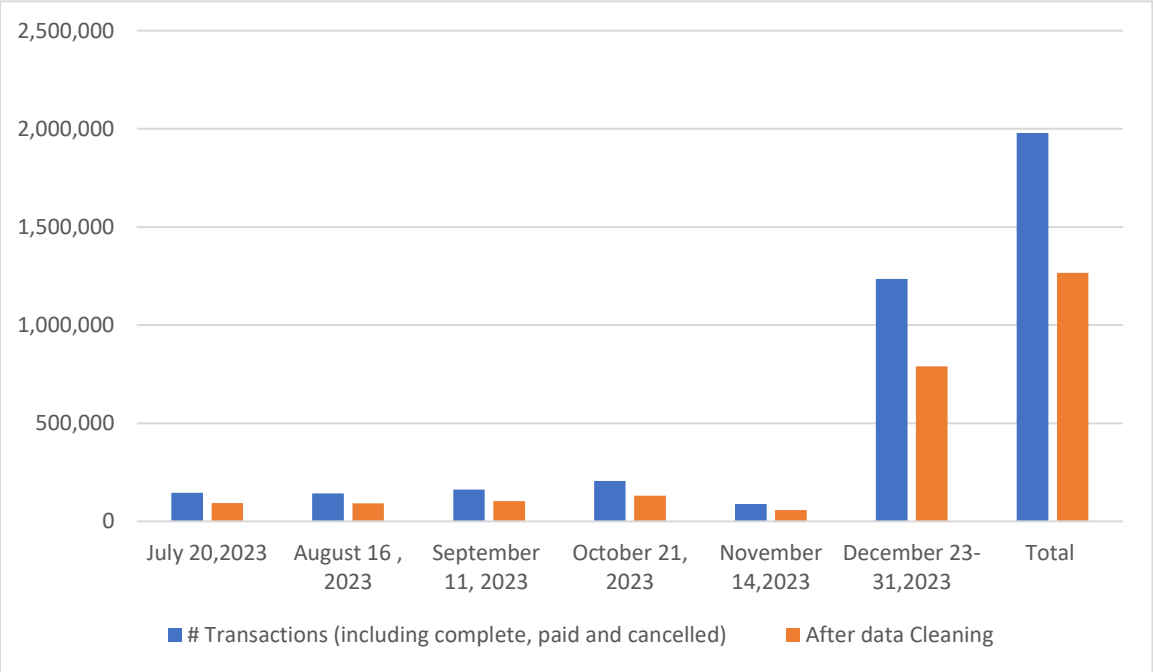
The columns are:

- TRANS_STATUS, TRANS_INITATE_TIME, TRANS_END_TIME TRANS_INITATE_TIME,
- TRANS_END_TIME, , DEBIT_PARTY_TYPE, DEBIT_ACCOUNT_TYPE
- CREDIT_PARTY_MNEMONIC, EXPIRED_TIME, , REQUEST_CURRENCY, EXCHANGE_RATE,
- CREDIT_PARTY_MNEMONIC,EXPIRED_TIME,REQUEST_AMOUNT, REQUEST_CURRENCY,
- EXCHANGE_RATE ACCOUNT_UNIT_TYPE, CURRENCY

The descriptive assessment of the dataset aimed to provide a comprehensive understanding of its overall structure and identify any potential issues, such as missing values or zero values, before conducting further analyses. Final data cleaning procedure goes as followings:

- Missing Values handling:

  Missing values in a dataset can pose challenges and may lead to inaccurate or biased results if not handled properly. In this case, any rows containing missing values were eliminated from the dataset. Out of the total 1,978,200 transactions initially retrieved from the database 1,266,048 transactions remained after removing the rows with missing values. This indicates that 712,152 transaction rows had missing values and were found therefore, excluded from the analysis. By eliminating the rows with missing values, the dataset was streamlined and made more suitable for subsequent analyses.



- Assess Missing Values:
  - The missing values are because of transaction has been cancelled, error, transaction interruption etc... As described by the business owners of the bank. The data have contained transactions with missing information like partial amounts. Cleaning removes such entries to ensure complete data for analysis.

- Remove Rows or Columns:

    - Column data removed are already stated above (i.e. columns are Removed as irrelevant: If a column has a large number of missing values or isn't relevant to the analysis, and such kind of records are moved)
- Zero Values:
    - Zero values are not meaningful in this case for fraud detection; it is treated them as missing values and apply imputation methods accordingly.
- Cancelled Transactions:
    - Cancelled transactions have been excluded during cleaning since they were not relevant for analysis purposes (e.g., not completed purchases or net worktime out or any other reason as explained from the bank staffs).
- Invalid or Erroneous Data:
    - Transactions with nonsensical data points, illogical values (e.g., negative transaction amounts), or invalid characters have been removed during cleaning as also impacts analysis. This also includes values that are not appropriate in the amount column after product name column data extraction.

The pre-processing, cleaning, and revoking of unnecessary fields from the general data set of mobile money transactions has been done removing tall the data points listed above. It is also noted that the significant difference in transaction volume for December might be because the data captured is a week of data.

4.5.2. Standardizing the phone number format (for mobile money data set):

From DEBIT_PARTY_MNEMONIC and CREDIT_PARTY_MNEMONIC column dataset contained customer phone numbers written in three different formats: 911xxxxxx, 0911xxxxxx, and 251911xxxxxx in addition customer. Having multiple formats for phone numbers can introduce inconsistencies. To address this issue and ensure data integrity, all phone number formats were standardized to 0911xxxxxx, which was the most common format among the phone numbers in the dataset. By changing the phone number formats to 0911xxxxxx, 400,387 phone numbers were modified. Standardizing the phone number format helps to maintain consistency and facilitates data analysis and support investigation when study is applied practically. This data could help on further analysis and investigation by human.

```python
import pandas as pd

def standardize_phone_numbers(df, column_name):
  """Standardizes phone numbers in a DataFrame column to the format 0911xxxxxx.

  Args:
    df: The pandas DataFrame containing the phone number column.
    column_name: The name of the column containing phone numbers.

  Returns:
    A new DataFrame with standardized phone numbers.
  """

  def standardize_phone_number(phone_number):
    if phone_number.startswith('911'):
      return f'0{phone_number}'
    elif phone_number.startswith('0911'):
      return phone_number
    elif phone_number.startswith('251911'):
      return phone_number[3:]
    else:
      return None  # Handle invalid phone numbers

  df[column_name] = df[column_name].apply(standardize_phone_number)
  return df

# data is in a CSV file named 'transactions1.csv'
df = pd.read_csv('transactions1.csv')

# Standardize phone numbers in the 'DEBIT_PARTY_MNEMONIC' column
df = standardize_phone_numbers(df, 'DEBIT_PARTY_MNEMONIC')

# Save the modified DataFrame to a new CSV file
df.to_csv('standardized_transactions.csv', index=False)
```

Figure 15 to extract standard phone number in 091xxxxxxx

4.6. Data Pre-processing:

From the mobile banking database, extracting relevant details from various fields, converting data types, filtering specific rows and columns, and restructuring tables for subsequent analysis. For instance, the SELECTED_OPTIONS column contained critical information such as the recipient's phone number and the transaction amount, which were extracted using Python regular expressions, as detailed in Appendix B.

The data preparation and pre-processing steps were performed multiple times to ensure the dataset was optimal for machine learning activities. As discussed in Section 3.1, the PRODUCT_NAME column was particularly important in categorizing transactions as Top-up, CBE-Birr, or Tele-Birr. Meanwhile, the SELECTED_OPTIONS column provided detailed transactional information, such as recipient details and amounts, further refined using Python regular expressions, as shown in the appendix.

4.6.1. Class Imbalance Problem in mobile banking and mobile money.

In this analysis, the problem of imbalance will be handled among the classes, and the class imbalance in the dataset will be assessed. The class imbalance is a percentage of the total number of transactions presented as fraudulent transactions that are reported by the bank as fraud.

As the case shared by the bank through the transaction reference numbers to keep the secrecy of the data. The author of this study has 2000 fraud or suspicious transactions mobile banking cases as reported from customer to the bank and bank run an investigation accept as frauds in the last one year. On the other had the mobile money side the transactions reported by the as fraud are 1150 in number.

Calculate the ratio of non-fraudulent transactions to fraudulent transactions for both mobile banking and mobile money. This gives a quantitative measure of the imbalance as shown below. Confirming that there is a typical class imbalance. As a balanced dataset would ideally have a ratio close to 1:1 which in these particular cases is different.

Table 7 class imbalance between normal transaction and the reported fraud cases

Channel	Reported Fraud case	Total transaction (sample transaction)	Ratio
Mobile banking	2000	24,131,109	8.28806E-05
Mobile money	1150	1,978,200	0.000581337

The results were found to be almost zero, indicating that there is a clear class imbalance in the cases under consideration here. Therefore, depending on the severity of the imbalance, various techniques can be applied, and in this study, the following techniques under sampling are employed: (Haibo He, and Edwardo A. 2009)

- Oversampling: Duplicating instances from the minority class to create a more balanced dataset.
- Under sampling randomly removing instances from the majority class to achieve balance.
- SMOTE (Synthetic Minority Oversampling Technique): Creates synthetic data points for the minority class to balance the dataset.
- Cost-sensitive learning: Assigns higher weights to misclassifications of the minority class during model training.

4.6.2. Rationale Behind Why the Under Sampling Chosen

As the data availability of the suspicious transactions are much less than the number of transactions, the dataset is already limited, for the minority class, oversampling might not be feasible here. There may not be enough data points in the minority class to generate a sufficient number of synthetic samples using techniques like SMOTE without introducing overfitting. Oversampling can amplify existing noise in the majority class data. Under sampling on the other hand could help remove some of this noise, focusing the SVM on learning the true patterns from the minority class. In addition, interpretability is another point in selection of the under sampling, which can help maintain the interpretability of the SVM model to some extent. This is it is understandable why the model makes certain predictions, which might be crucial for tasks like fraud detection. In addition, the under sampling is less computation-intensive to run.

4.6.3. Experiment for the under sampling to overcome the class imbalance problem:

1. Understand the Class Imbalance:

- In class distribution, the minority class are the reported fraud cases and majority class in this study are the randomly selected transactional data. This has been identified the calculations are done as show in table 9 above. In fraud detection, datasets often exhibit severe class imbalance, with a disproportionately low number of fraudulent cases compared to legitimate transactions, as represented in the study. This imbalance can significantly impact model performance, leading to models that are good at predicting the majority class (legitimate transactions) but poor at detecting the minority class (fraudulent transactions).

2. Data Splitting:

- The original dataset is split into training and testing sets to ensure that the evaluation of the model is done on unseen data. This is done using the following procedure:
    - Data Splitting:
        - The dataset is initially split into training and testing sets. This ensures that feature selection is performed only on the training data to prevent data leakage.
    - Feature Selection:
        - Feature selection is performed on the training set. This way, the test set remains untouched and unbiased, providing a reliable evaluation of the model.

3. Addressing the Class Imbalance:
    - To address the severe class imbalance in the dataset, which has 2,000 labeled fraud cases and 24 million normal transactions, undersampling is employed to balance the dataset before splitting it for model training.
    - The normal transaction data is collected into a NormalTransaction.csv file, and the labeled fraud data is collected into a FraudData.csv file using the provided Python script below.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from imblearn.under_sampling import RandomUnderSampler

# Load labeled fraud data
fraud_data = pd.read_csv("labledFraudData.csv.csv")

# Load normal transactions data
normal_data = pd.read_csv("NormalTransaction.csv")

```

Figure 16 Merging the labelled and unlabelled data using the under sampling

4.6.4. Under sampling the Majority Class

The RandomUnderSampler used in these cases to reduce the majority class (normal transactions) to match the size of the minority class (fraud cases) for both mobile banking and mobile money data set.

Balancing the Dataset: Ensures that the training dataset has an equal number of fraud and normal transactions, which is crucial for training models effectively in fraud detection scenarios for this the following code is applied on the merged data set.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from imblearn.under_sampling import RandomUnderSampler

# Load labeled fraud data
fraud_data = pd.read_csv("labledFraudData.csv")

# Load normal transactions data
normal_data = pd.read_csv("NormalTransaction.csv")

# Combine fraud and a random sample of normal transactions
normal_sample = normal_data.sample(n=len(fraud_data), random_state=42)
combined_data = pd.concat([fraud_data, normal_sample], axis=0)

# Apply RandomUnderSampler
rus = RandomUnderSampler(random_state=42)
X_resampled, y_resampled = rus.fit_resample(combined_data.drop('Fraud', axis=1), combined_data['Fraud'])
```

a. Sampling Normal Transactions:
    a. normal_data.sample(n=len(fraud_data), random_state=42): Randomly selects a subset of normal transactions equal in size to the number of fraud transactions. The random_state=42 ensures reproducibility of the sample.
b. Combining Data:
    a. pd.concat([fraud_data, normal_sample], axis=0): Concatenates the fraud transactions and the sampled normal transactions into a single DataFrame, combined_data. This balances the dataset by having an equal number of fraud and normal transactions.

c. Applying RandomUnderSampler:
    a. RandomUnderSampler(random_state=42): Initializes the RandomUnderSampler, which is used to further balance the dataset by undersampling.
    b. rus.fit_resample(combined_data.drop('Fraud', axis=1), combined_data['Fraud']):
        i. combined_data.drop('Fraud', axis=1): Selects all columns except the target variable 'Fraud' for features.
        ii. combined_data['Fraud']: Selects the 'Fraud' column as the target variable.
        iii. fit_resample: Resamples the dataset to balance the classes based on the target variable. This step ensures that both classes (fraud and non-fraud) are equally represented after resampling.

The outcome was the combination of sampled data: non-fraudulent transactions were merged with labelled fraudulent transactions to create a new balanced dataset.

- **Combined Dataset resulted in for th mobile banking: 4,000 total records (2,000 fraud + 2,000 non-fraud)**
- **Combined data set resulted in for the mobile money data set total records (1150 fraud + 1150 non –fraud)**

4.6.5. How the data splitting applied:

A stratified train-test split is applied in this experiment in that it is a method of splitting a dataset into training and testing subsets in such a way that the class proportions (or distributions of the target variable) are the same in both subsets as they are in the overall dataset. A stratified split ensures that the training and testing sets maintain the same proportion of each class as the original dataset. This is particularly important when dealing with imbalanced data situation like in this study, where one class (e.g., normal transactions) significantly outweighs another (e.g., fraudulent transactions).

4.7.Modelling Training and Evaluation

4.7.1. Selection of SVM-based Fraud Detection Model

In this section, the study explores the rationale behind selecting a Support Vector Machine (SVM) as the primary algorithm for fraud detection and training and evaluation also be discussed. SVM is renowned for its ability to handle large feature spaces and deliver robust classification results, making it particularly suitable for fraud detection where the distinction between fraudulent and legitimate transactions may be subtle (Vapnik, 1998). The selection

process involves carefully considering feature importance, where the transaction amount plays a pivotal role as an indicator of potential fraud in this study as already explained.

4.7.2. Training the Model

Training the SVM model involves feeding it with pre-processed transaction data, where features such as "Amount" are standardized to ensure uniformity in the learning process (Hastie, Tibshirani, & Friedman, 2009). A time series analysis is conducted as part of the feature engineering process, capturing trends, seasonality, and potential outliers that may indicate fraudulent activities (Hyndman & Athanasopoulos, 2018). The training process is iterative, where the model learns to differentiate between fraudulent and non-fraudulent

```
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.metrics import classification_report, roc_auc_score, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
from statsmodels.tsa.seasonal import seasonal_decompose

# Convert date fields to datetime
df['PROCESSED_DATE'] = pd.to_datetime(df['PROCESSED_DATE'])
df['AUTH_CODE_EXPIRY_DATE'] = pd.to_datetime(df['AUTH_CODE_EXPIRY_DATE'])

# Extract additional features from the date
df['PROCESSED_YEAR'] = df['PROCESSED_DATE'].dt.year
df['PROCESSED_MONTH'] = df['PROCESSED_DATE'].dt.month
df['PROCESSED_DAY'] = df['PROCESSED_DATE'].dt.day
df['PROCESSED_DAYOFWEEK'] = df['PROCESSED_DATE'].dt.dayofweek

# Handling the Amount feature
amount_feature = ['Amount']

# Encoding the Amount_Binned feature
if df['Amount_Binned'].dtype == 'object':
    le = LabelEncoder()
    df['Amount_Binned_Encoded'] = le.fit_transform(df['Amount_Binned'])
    amount_feature.append('Amount_Binned_Encoded')

# Scaling the Amount
scaler = StandardScaler()
df['Amount_Scaled'] = scaler.fit_transform(df[['Amount']])
amount_feature.append('Amount_Scaled')

# Preparing the feature set X and target y
X = df[amount_feature + ['PROCESSED_YEAR', 'PROCESSED_MONTH', 'PROCESSED_DAY', 'PROCESSED_DAYOFWEEK']]
y = df['Target']
```

transactions through a series of optimization steps (Schölkopf & Smola, 2002). The python script used for time series analysis, as outlined below, includes algorithms for detecting patterns that are characteristic of fraudulent behavior over time.

```
# Splitting the data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_s

# Training the SVM Model
model = SVC(kernel='rbf', probability=True)
model.fit(X_train, y_train)

# Predictions
y_pred = model.predict(X_test)
y_pred_prob = model.predict_proba(X_test)[:, 1]

# Evaluation
classification_rep = classification_report(y_test, y_pred)
roc_auc = roc_auc_score(y_test, y_pred_prob)

# Confusion Matrix
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(10, 7))
sns.heatmap(cm, annot=True, fmt='d', cmap="Blues")
plt.title('Confusion Matrix')
plt.show()

# Time series analysis
df.set_index('PROCESSED_DATE', inplace=True)

# Plotting transaction counts over time
plt.figure(figsize=(12, 6))
daily_fraud = df['Target'].resample('D').sum()
daily_fraud.plot(title='Daily Fraudulent Transactions')
plt.show()

# Aggregating data by month for a broader view
plt.figure(figsize=(12, 6))
monthly_fraud = df['Target'].resample('M').sum()
monthly_fraud.plot(title='Monthly Fraudulent Transactions')
plt.show()

# Identifying any anomalies or seasonal patterns
result = seasonal_decompose(df['Target'].resample('D').sum(), model='additive')
result_plot = result.plot()
plt.show()

# Print results
classification_rep, roc_auc
```

Figure 17 Model training and evaluation Python script from mobile banking data

4.7.3. Data Pre-processing and Feature Engineering Experiment

Before training the model, the dataset underwent essential pre-processing steps (where details explained in the data preparation section) few of the main experiments are:

1. **Date Conversion:**
    - The PROCESSED_DATE and AUTH_CODE_EXPIRY_DATE fields were converted to datetime objects. This allowed to extract temporal features like the year, month, day, and day of the week, which were hypothesized to contribute to identifying fraudulent patterns.
2. **Feature Extraction:**
    - From the PROCESSED_DATE, features such as PROCESSED_YEAR, PROCESSED_MONTH, PROCESSED_DAY, and PROCESSED_DAYOFWEEK were extracted. These features aimed to capture potential temporal patterns related to fraud.
3. **Handling of the Amount Feature:**
    - The Amount feature was scaled using standardization to ensure that the SVM model, which is sensitive to the scale of input features, could effectively process the data.
    - The Amount_Binned feature, representing categorical bins of transaction amounts (e.g., Low, Medium, High), was label-encoded to convert it into a numerical format. This allowed the model to consider different transaction amount categories in its decision-making process.

4.8. Results Discussion Up On the Execution of Experiment

4.8.1. Analysis of the SVM Model Results

1. **Confusion Matrix**:

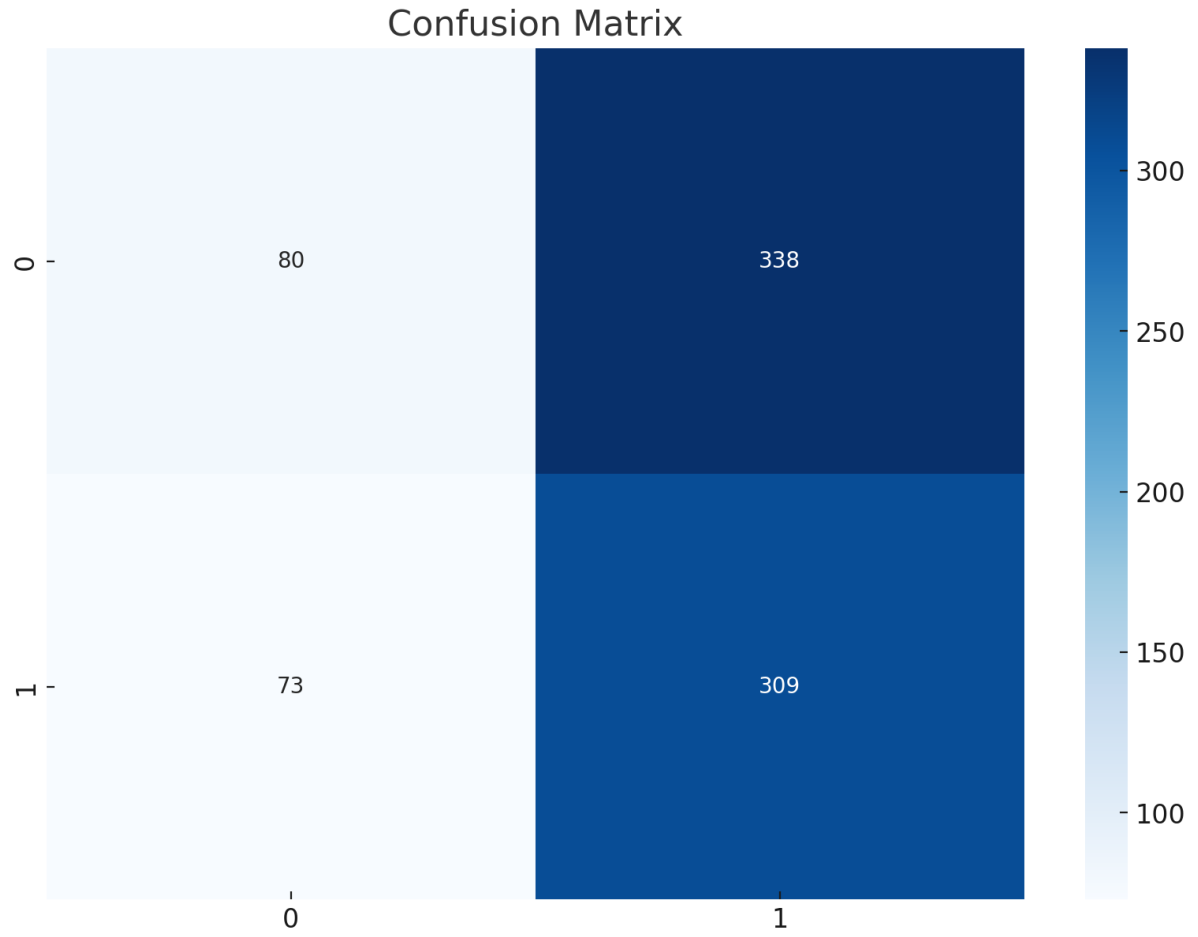


Figure 18 Confusion Matix

	Predicted: Non-Fraud (0)	**Predicted: Fraud (1)**
Actual: Non-Fraud (0)	True Negatives (TN)	False Positive (FP)
Actual: Fraud (1)	False Negatives (FN)	True Positives (TP)

The confusion matrix shows how the SVM model performed in predicting fraudulent (label 1) and non-fraudulent (label 0) transactions. The matrix indicates that the model struggled to correctly classify non-fraudulent transactions, as reflected by a high number of false positives. Conversely, the model correctly identified many fraudulent transactions but also missed some.

2. **Classification Report (Based on the confusion matrix):**

- Precision for Non-Fraudulent Transactions (0): 0.52 - This means that when the model predicts a transaction as non-fraudulent, it is correct 52% of the time.

Calculation➔ Precision=True Positives / (False Positives +True Positives)

- Recall for Non-Fraudulent Transactions (0): 0.19 - This low recall indicates that the model only correctly identifies 19% of the actual non-fraudulent transactions, leading to many false positives.

Calculation➔Recall=True Positives / (False Negatives + True Positives)

- Precision for Fraudulent Transactions (1): 0.48 - For fraudulent transactions, the precision is slightly lower, at 48%.
- Recall for Fraudulent Transactions (1): 0.81 - The model correctly identifies 81% of fraudulent transactions, showing better performance in fraud detection.
  Overall Accuracy: 0.49 - The overall accuracy is around 49%, which is not particularly strong and suggests room for model improvement.

  Calculation➔F1-Score=2× ((Precision+Recall)/(Precision×Recall)
  - The F1-score is the weighted average of precision and recall. It provides a balance between the two, especially when dealing with imbalanced classes.
- The **ROC-AUC score** (Receiver Operating Characteristic - Area Under the Curve) is a performance metric for evaluating the ability of a binary classification model to distinguish between classes (in this case, fraudulent and non-fraudulent transactions). The ROC-AUC score of **0.51** is slightly above 0.5, indicating that the model's performance is only marginally better than random guessing.

1. **Time Series Analysis**

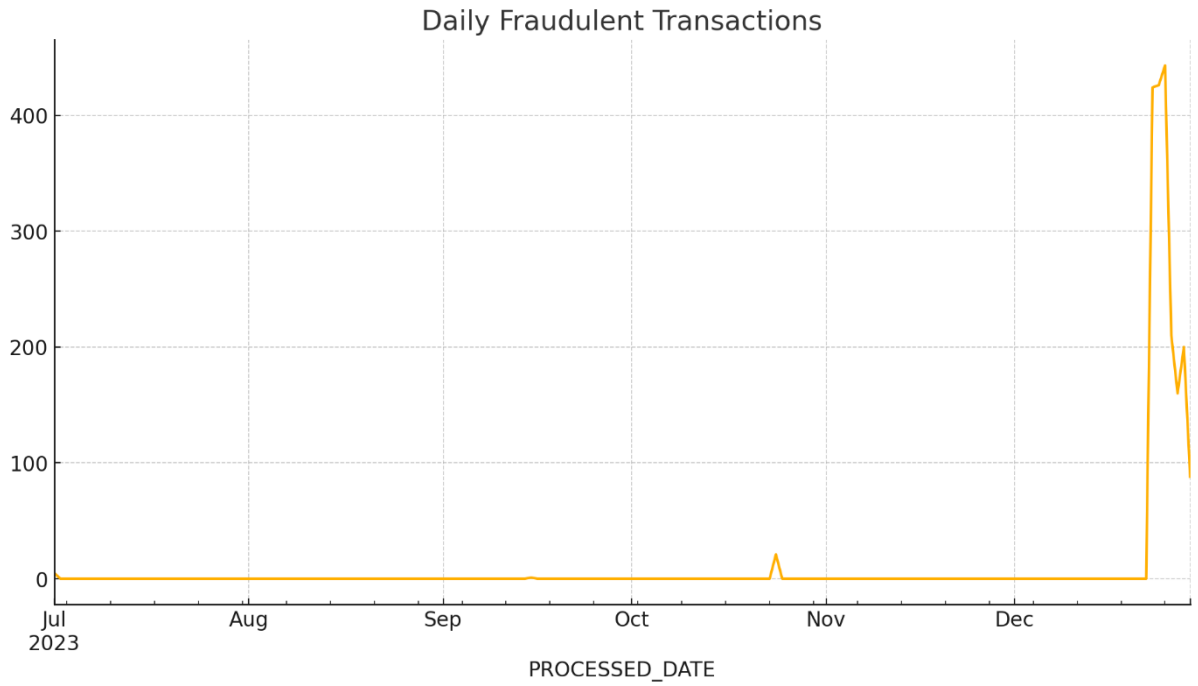


Figure 19 Time Series Analysis Daily Fraudulent Transactions

Daily Fraudulent Transactions: The daily time series plot of fraudulent transactions shows fluctuations in fraud activity over time. Peaks in the plot might correspond to specific events or periods when fraudulent activity increased in addtion the data et was taken more transaction from the December transactions. This kind of analysis is crucial for identifying trends that could inform model adjustments or additional preventative measures. However, as the data is just only from the fraud cases reported and the number of reports investigated are depending on the team size this could may show the fraud trends over time are similar that the actual happening.

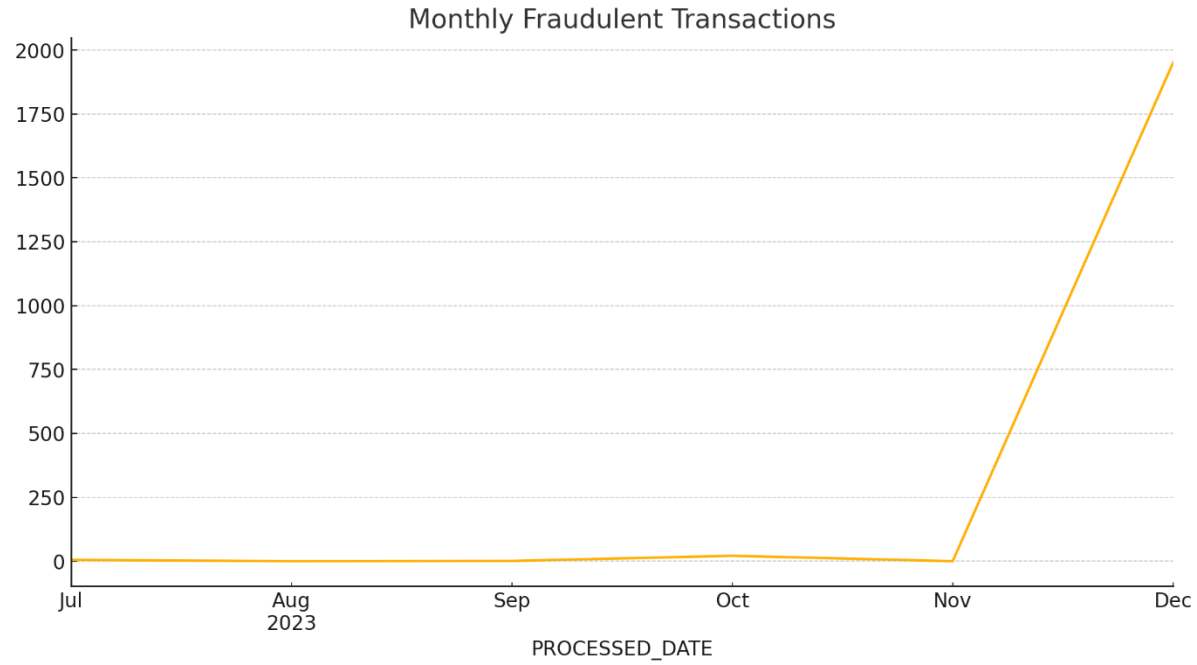


Figure 20 Time Series Analysis Monthly Fraudulent Transactions

Monthly Fraudulent Transactions: Aggregating data monthly provides a broader perspective, revealing any long-term trends or seasonal patterns in fraudulent transactions. The patterns cold suggest that certain times of the year are more prone to fraud, suggesting targeted intervention strategies but fraud or any weakness on systems. In addition as discussion with the bank experts said that transaction will yearly spick starting October as credit disbursement for all type of business is released so as fraudulent transactions.

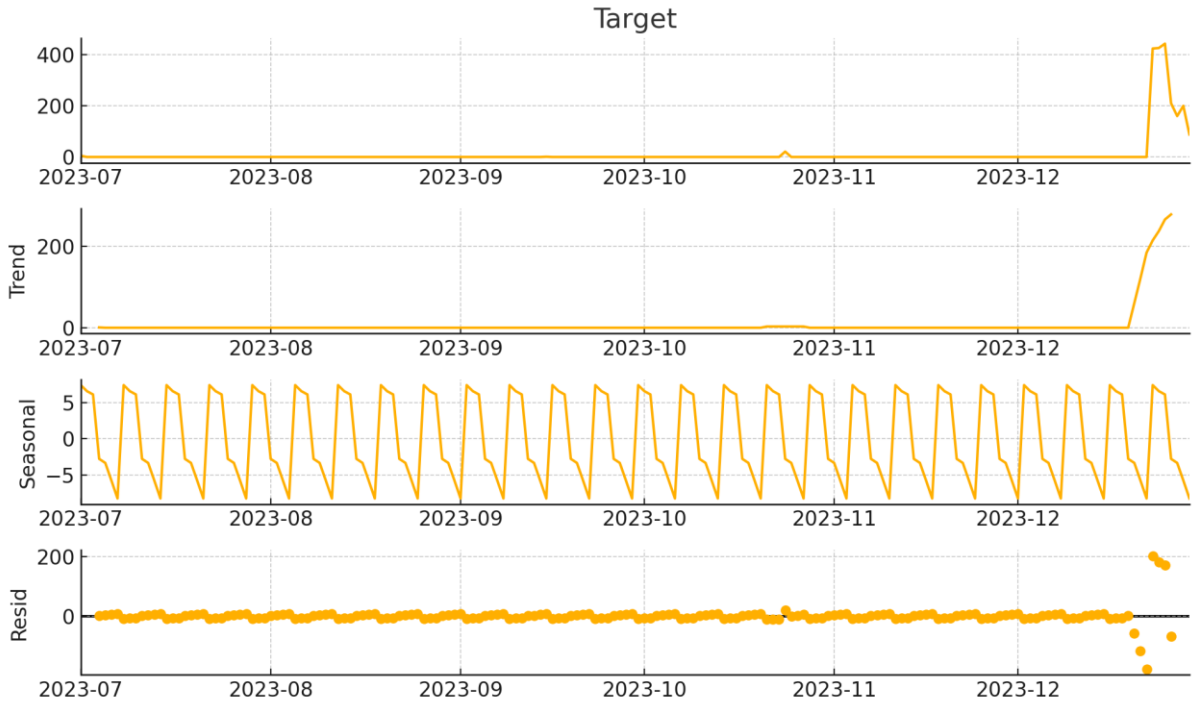


Figure 21 Seasonal Decomposition

4.8.2. Conclusion SVM model's performance

The SVM model's performance on dataset indicates a higher recall for fraudulent transactions but struggles with precision and overall accuracy. The time series analysis underscores the importance of understanding fraud patterns over time, which could be leveraged to refine the model or develop complementary approaches to fraud detection in addition to how transactions are logged in the source systems of the bank. The model's limitations suggest the need for further feature engineering, model tuning, or even exploring alternative models to improve its performance.

4.8.3. Evaluation on a real cases

The accuracy of an SVM model after training is evaluated through the following procedures by the bank experts:

The model was applied to the entire dataset, including both mobile banking and mobile money transactions. In the mobile money channel, the model predicted 600 cases as

fraudulent. This result was submitted to the bank for further investigation by four business analysts.

Each analyst was tasked with investigating a subset of these cases, covering a total of 89 transactions (29 + 40 + 15 + 5). Out of these, the analysts confirmed that 19 cases were indeed fraudulent, while the remaining transactions were determined to be genuine.

Regarding the mobile money transactions, a bank analyst provided an oral report indicating that they identified some cases that appeared suspicious. However, they were unable to definitively classify these cases as fraudulent putting a reason that they can't confirm from customers about the flagged cases as fraud.

4.8.4. Deployment

In this study, the author explores the feasibility of detecting fraud in mobile banking transactions using supervised machine learning techniques, specifically focusing on Support Vector Machine (SVM) models. The primary goal is to determine whether these techniques can effectively identify fraudulent activities the below shows how the process goes and deployed.

Deployment Process

This deployment framework demonstrates the practical application of SVM for detecting fraud in mobile banking while establishing a foundation for ongoing improvements and adaptations to meet emerging fraud patterns. With the framework overview, model evaluation, and implementation processes discussed, the brief implementation how it works is as follows:

1. **Data Collection and Cleaning:** Transaction data is collected and cleaned in Google Colab. For instance, missing transaction values are imputed, and features are normalized. In this case, a labelled dataset of 2,000 mobile banking transactions and 1,150 mobile money transactions was randomly selected. This data then merged with a randomly selected portion of unlabelled data, and the stratified train-test split is applied for 80% training and 20% test cases while selecting the transaction amount as a feature.
2. **Training the Model:** The SVM model is trained on based on a spited rom the above step to let learn patterns by the SVM algorithm associated with fraudulent transactions.

3. **Evaluating Performance:** The model's performance is evaluated using metrics like precision and recall to ensure it accurately detects fraud while minimizing false positives. In this particular cases transaction classification is as fraud and genius has been a struggle for the the model as outlined in the evaluation section of this study.
4. **Deployment:** The trained model is deployed in a data set where transactions are stored for analysis analysed as they captured while randomly groping into 4000 cases for mobile banking and another . The model flags suspicious transactions for further investigation which a single execution of the cases to the bank experts for testing. In their response was

   "The model was applied to the entire dataset, including both mobile banking and mobile money transactions. In the mobile money channel, the model predicted 600 cases as fraudulent. Each analyst was tasked with investigating a subset of these cases, covering a total of 89 transactions (29 + 40 + 15 + 5). Out of these, the analysts confirmed that 19 cases were indeed fraudulent, while the remaining transactions were determined to be genuine. In the cases of mobile banking.

5. **Monitoring and Updating:** in this cases data getting data in real time through API for fraudulent classification is impossible in this study. However when the cases is acceptable form the banks to apply detection either in real time or near real time, the model could be integrated on a locally set up enviroment to classify transactions as fraud.

CHAPTER FIVE

5. Conclusion and Recommendation

5.1. Conclusion

This study set out to assess the effectiveness of machine learning (ML) techniques in detecting fraudulent transactions in the Ethiopian mobile banking and mobile money sectors. With the rapid growth of mobile financial services in Ethiopia, the threat of fraud has become a significant concern. Addressing this issue through advanced technology is crucial for maintaining trust in the financial system. The study's findings provide valuable insights into the capabilities and limitations of ML models in this context, particularly considering the challenges posed by data quality and availability.

The study contributes to the understanding of various fraud patterns, such as transactions occurring immediately after profile changes, suspicious mobile top-ups orchestrated by fraudsters, or accounts receiving funds from multiple sources. The tested Support Vector Machine (SVM) model successfully detected over 50% of these cases. However, despite the model's effectiveness, there remains a critical need for manual investigation to confirm and classify these detected incidents as fraud.

Despite the class imbalance in the dataset, SVM was utilized in this study. While SVM is not inherently robust to class imbalances, it was chosen for its effectiveness in high-dimensional spaces and the availability of techniques, such as adjusting class weights, to mitigate the impact of imbalance. Although better results for detecting fraud can be achieved with this development, there still remains a need for manual investigations into each detection precision is not as expected.

The analysis of findings indicates the presence of several suspicious transactions that require manual investigations to be classified as fraud by analysts after transactions are detected by machine learning technique, such as the one explored in this study.

Moreover, this study highlights the potential of ML to discover unreported fraud incidents and detect suspicious behaviour among high-risk customers. As financial institutions increasingly adopt digital banking innovations, they must also implement robust fraud

management systems that leverage ML and data analytics, as out of thousands of transactions detection suspicious once is almost impossible to humans. These systems are vital in the growing digital banking landscape, enabling the investigation and prevention of significant fraudulent activities that could have severe consequences for the entire country.

Effectiveness of Machine Learning Models in Fraud Detection:

The first research question aimed to assess the effectiveness of ML models in detecting fraud within the Ethiopian banking sector, where data quality may be limited. The study utilized a Support Vector Machine (SVM) model to evaluate its performance on a dataset derived from mobile banking and mobile money transactions irrespective of the labled data quality challenge. While the model showed promise in detecting fraudulent activities, the overall performance was suboptimal, particularly in identifying fraud cases accurately.

The SVM model's ROC AUC score, which was approximately 0.51, indicated that the model's ability to distinguish between fraudulent and non-fraudulent transactions was marginally better than random guessing. The confusion matrix further highlighted the model's limitations, as it failed to correctly classify any fraudulent transactions, resulting in a high rate of false negatives. This outcome underscores the challenge of applying ML techniques in an environment where data quality and volume is a challenge but still need testing various scenarios of ML to apply.

Addressing data Imbalances:

The second research question focused on how data imbalances could be addressed to optimize ML model performance for fraud detection in the Ethiopian context. Data imbalance is a common issue in fraud detection, where fraudulent transactions typically represent a small fraction of the overall dataset. This imbalance can lead to models that are biased toward predicting the majority class (non-fraudulent transactions), as observed in this study.

To address this, under sampling is utilized as being and easy and less computationally expensive to apply. Additionally, experimenting with different ML algorithms that are less sensitive to class imbalances, such as Random Forest could potentially improve model

performance. However, these approaches would require careful consideration of the Ethiopian banking sector's specific data characteristics and the computational resources available.

General and Specific Objectives Achievement:

The general objective of the research was to quantitatively assess the effectiveness of ML techniques for detecting fraud in Ethiopian mobile banking and mobile money services. The study successfully explored this objective by applying a ML technique (SVM) on a relevant dataset. Although the model's performance highlighted challenges, the research provided he capabilities and constraints of ML in this context as using the less fraud labelled impacts the result as the inclusion of synthetic data may impact fraud detection. This requests the finance sector to seriously consider the investigation and recording of fraud for future application of ML.

5.2. Recommendations

After reviewing the literature and conducting data analysis in this study, several recommendations are offered to the digital banking industry for the application of data analytics in fraud management.

Enhanced Data Collection and Quality Improvement:

The effectiveness of ML models is highly dependent on the quality and quantity of data. Ethiopian banks should invest in better data collection and management systems to ensure that transactional data is accurate, complete, and timely. This will provide a more reliable foundation for ML-based fraud detection.

Implement regular audits of data to identify and rectify issues related to data quality, such as missing values, inconsistencies, and errors. This will help improve the reliability of ML models.

Experiment with other ML algorithms to handle imbalanced datasets, such as Random Forest, or deep learning approaches. These models might offer better performance in detecting rare fraudulent transactions.

Encouraging collaboration between banks for sharing anonymized fraud data could lead to the development of more robust ML models that are better equipped to detect sophisticated fraud schemes.

In additions, it is noted that relying solely on a single data analytics technique may not always yield the most accurate fraud detection solution, as highlighted in the article by Deepa and Lalita (2023). Combining insights from the network structure between different customer accounts with the predictive power of supervised algorithms can lead to highly precise fraud detection. Therefore, leveraging machine-learning capabilities in this context could significantly contribute to combating fraud, particularly with the expansion of digitization.

Furthermore, transaction data analytics should encompass user behaviors, an aspect not explored in this study but one that could enhance the fight against fraud when combined with analyses conducted across various channels such as debit cards, credit cards, and other banking services.

As a third recommendation, implementing proactive mechanisms such as educating individuals about the potential risks of fraud and employing multifactor authentication to authorize customer transactions is suggested.

This study demonstrated the possibility of accurately identifying fraudulent transactions in financial data, despite a significant class imbalance. However, addressing unreported fraud cases within banks poses a considerable challenge for research and development efforts in the field of fraud detection.

➔End ➔

6. REFERENCES

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York, NY: Springer.
2. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* (Vol. 112). Springer.
3. Golzar, Jawad & Tajik, Omid & Noor, Shagofah. (2022). Convenience Sampling. 1. 72-77. 10.22034/ijels.2022.162981.
4. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039
5. Japkowicz, N., & Stephen, S. (2002). The Class Imbalance Problem: A Survey. **Intelligent Data Analysis**, 6(6), 429-449. https://www.sciencedirect.com/science/article/abs/pii/S0020025519310497
6. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Whalen, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, 16, 321-357. https://arxiv.org/pdf/1106.1813
7. Ahmad, F., Imran, M., Kang, B., & Lee, Y. (2023). Machine Learning for Network Anomaly Detection: A Survey. **Sensors (Switzerland)**, 23(2), 723. https://www.mdpi.com/1424-8220/23/3/1352
8. Ahmad, F., Imran, M., Kang, B., & Lee, Y. (2023). Machine Learning for Network Anomaly Detection: A Survey. **Sensors (Switzerland)**, 23(2), 723. https://www.mdpi.com/1424-8220/23/3/1352
9. Ghosh, S., & Roy, A. (2012). Fraud Detection in Mobile Banking using Data Mining Techniques. **International Journal of Computer Science & Information Technology (IJCSIT)**, 4(2), 3747.
10. Singh, A., Sharma, N., & Iqbal, R. (2023). A Comprehensive Survey on Data Imbalance Problem: Techniques, Applications and Evaluation Metrics. **ACM Computing Surveys (CSUR)**, 55(3), 1-40.
11. Kou, Y., Lu, Y., & Xue, Y. (2020, December). Real-time fraud detection system for e-commerce based on machine learning. In 2020 International Conference on Artificial Intelligence and Computer Science (AICS) (pp. 1072-1077). IEEE.
12. Phua, C., Lee, V., & Gayen, K. (2010, October). A comprehensive survey on intrusion detection systems (IDS). ACM Computing Surveys (CSUR), 42(4), 1-42.
13. Alaiwa, M., Almseidin, M., & Hajjaty, S. (2022). Machine learning and deep learning for intelligent fraud detection: A survey. *Journal of Big Data*, 10(1), 1-42. https://ieeexplore.ieee.org/document/9004231
14. Luo, F.-L., & Liu, H. (2017). Support Vector Machines for Fraud Detection in Credit Card Transactions. [Journal name], [volume number], [page range].

15. Decker, P. March/April 1998. Data Mining's Hidden Dangers. Banking Strategies, 6-14.
16. G. Jaculine Priya and Dr.S. Saradha, (2021),Fraud Detection and Prevention Using Machine Learning Algorithms: A Review
17. Oladimeji Kazeem, 2023, FRAUD DETECTION USING MACHINE LEARNING
1. G.Jaculine Priya and Dr.S.Saradha (2021),Fraud Detection and Prevention Using Machine Learning Algorithms: A Review
2. Reurink, Arjan. (2016). Financial fraud: a literature review. Cologne: Max Planck Institute for the Study of Societies.
3. PricewaterhouseCoopers (PwC). (2016). Global economic crime survey 2016. https://www.pwc.com/gx/en/economic-crime-survey/pdf/GlobalEconomicCrimeSurvey2016.pdf
4. Association of Certified Fraud Examiners (ACFE). (2022). Occupational fraud 2022: A report to the nations. https://legacy.acfe.com/report-to-the-nations/2022/
5. Haibo He, and Edwardo A. (2009), Garcia Learning from Imbalanced Data
6. National Bank of Ethiopia Quarterly Bulletin (2022/23)
7. Yonas Worku, (2020), FRAUD IN THREE ETHIOPIAN BANKS FOLLOW ON TECHNOLOGY DRIVEN BANKING SERVICES, Addis Ababa University Faculty of Business and Economics Department of Management
8. mlbookcamp (2024), Machine Learning Bookcamp [Online],[Accessed January 01, 2023), https://mlbookcamp.com/article/crisp-dm
9. Rüdiger Wirth (2017), CRISP-DM: Towards a Standard Process Model for Data, Mining
10. Martínez-Plumed, Fernando, et al.(2019) "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories." IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 8, 2019, pp. 1833-1843.
11. Frehiwot Mola (2019) Analysis and Detection Mechanisms of SIM Box Fraud in The Case of Ethio telecom, Addis Ababa university
12. Andrea Willige (November, 2023), Here's why Africa is the world leader in digital and mobile banking [Online] [Accessed January 14, 2024]]https://etradeforall.org/news/heres-why-africa-is-the-world-leader-in-digital-and-mobile-banking/
13. Daniel, M. (2013). Application of Datamining Technology to support to support fraud protection. Addis Ababa: Addis Ababa University.
14. Ali, A.; Abd Razak, S.; Othman, S.H.; Eisa, T.A.E.; Al-Dhaqm, A.; Nasser, M.; Elhassan, T.; Elshafie, H.; Saif, A. Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. Appl. Sci. 2022, 12, 9637. https://doi.org/10.3390/app12199637

15. Ayesha Karmustaji (July, 2021) Fraud Detection using Data Analytics, A Capstone Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Professional Studies, RIT Dubai
16. Daniel, M. (2013). Application of Datamining Technology to support to support fraud protection. Addis Ababa: Addis Abeba Univeristy.
17. Adane, T. (2011). Mining insurance data for fraud detection. Addis Ababa University, Addis Ababa, Ethiopia.
18. Tsegaye Nire(2017, Constructing a predictive model for Real-Time ATM CARD Fraud Detection (Master's thesis Addis Ababa university)
19. Kou, Y., Lu, C.-T., Sirwongwattana, S., & Huang, Y.-P. (2004). Survey of fraud detection techniques. IEEE International Conference on Networking, Sensing and Control, 2004. 2, pp. 749-754. Taipei, Taiwan: IEEE.
20. Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & 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.
21. Catherine Cote (December 2021), 7 Data Collection Methods in Business Analytics, [Online] [Accessed December 10, 2023], https://online.hbs.edu/blog/post/data-collection-methods
22. Deepa and Lalita (2023) A systematic literature review on frauds in banking sector, Deepa Mangala and Lalita Soni
23. Creswell**,** John W. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. SAGE Publications.
24. Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. Expert Systems with Applications, 193
25. Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. Computer Science Review, 40.
26. Ruchir Dahal (2021), Qualitative and Quantitative Data Analysis Methods, [Online], [Accessed December 08, 2023], https://www.grepsr.com/blog/qualitative-and-quantitative-data-analysis-methods/
27. Jarrod West, Maumita Bhattacharya, Intelligent financial fraud detection: a comprehensive review, Computers & Security (2015),
28. Ic3(2022), Criminals Increasing SIM Swap Schemes to Steal Millions of Dollars from US Public [Online] [Accessed 30th November, 2023] https://www.ic3.gov/Media/Y2022/PSA220208

29. Net Guardians (2023) The Top Banking Fraud Types to Watch in 2023 [Online] [Accessed 30 November 30th 2023] https://www.netguardians.ch/the-top-banking-fraud-to-watch-in-2023/#chapter215
30. Onespan (2023), What is account takeover fraud (ATO)? [online] [Accessed 26th November 2023] https://www.onespan.com/topics/account-takeover-fraud
31. Amy Lurie (2023) What is Anti Money Laundering (AML) and How Does It Work? *[online] [Accessed 25th November 2023]* https://www.au10tix.com/blog/anti-money-laundering-aml/
32. *Joseck Luminzu Mudiri (2013). Fraud in Mobile Financial Services [unpublished]. A MicroSave Publication,*
33. *Anna Sorbet. (2022) History of mobile banking – how it all started? [online] [Accessed on 25th November 2023]* https://finanteq.com/blog/fintech-trends/history-of-mobile-banking-how-it-all-started/
34. *Ali Abdallah Alalwan a, Yogesh K. Dwivedi b,⁎, Nripendra P. Rana b, (2017), Factors influencing adoption of mobile banking by Jordanian bank customers: Extending UTAUT2 with trust*
35. *Aijaz A. Shaikh, Heikki Karjaluoto, Mobile banking adoption: A literature review, Telematics and Informatics, Volume 32, Issue 1, 2015,*
36. *TechBullion. (n.d.).(2023) , What is a Mobile Wallet, Origin and History in Financial technology? - [Online] [Retrieved August 26, 2023], The impact of private sector projects in Africa - Studies from the EIB-GDN Programme, https://techbullion.com/mobile-wallet-origin-history-financial-technology/*
37. *Peacock, M. A. (2021). Telecommunications System & Management An Analysis of the Potential Risk and Fraud Involved in Mobile Money Transaction in Freetown Sierra Leone. 10, 10–13.*
38. Rezaee, Z., & Wang, J. (2019). Relevance of big data to forensic accounting practice and education. Managerial Auditing Journal, 34(3), 268–288. https://doi.org/10.1108/MAJ-08-2017-1633
39. Chuprina, Roman (13 April 2020). "The In-depth 2020 Guide to E-commerce Fraud Detection". www.datasciencecentral.com. Access Date 2023-11-14.
40. How safe is your money in banks? The reporter (Ethiopia Magazine) https://www.thereporterethiopia.com/25850/#google_vignette (Access date November 2023)
41. Fraud management: Recovering value through next-generation solutions, (August 20, 2018) By Lindsay Anan, Robert Hayden, Kaustubh Joshi, Marie-Claude Nadeau, and Jonathan Steitz, Accessed date: (November 2023) https://www.mckinsey.com/industries/financial-services/our-insights/fraud-management-recovering-value-through-next-generation-solutions

42. Salim, Rob, and Rob (Mckinsey article) https://www.mckinsey.com/industries/financial-services/our-insights/combating-payments-fraud-and-enhancing-customer-experience#/ access date(10 November 2023)
43. Abiy Woretaw and Lemma Lessa (2012). Information security culture in the banking sector in Ethiopia. https://docplayer.net/12528591-Information-security-culture-in-thebanking-sector-in-ethiopia.html (Accessed date: November 2023)

7. ANNEX

```
SELECT *
FROM cps_transaction
WHERE transaction_date >= TO_DATE('2023-07-01', 'YYYY-MM-DD')
  AND transaction_date <= TO_DATE('2023-08-07', 'YYYY-MM-DD')
   OR transaction_date = TO_DATE('2023-09-15', 'YYYY-MM-DD')
   OR transaction_date = TO_DATE('2023-10-24', 'YYYY-MM-DD')
   OR transaction_date = TO_DATE('2023-11-29', 'YYYY-MM-DD')
   OR (transaction_date >= TO_DATE('2023-12-23', 'YYYY-MM-DD') AND transactio
n_date <= TO_DATE('2023-12-31', 'YYYY-MM-DD'));
```

Annex 1 Mobile money data collection SQL query on oracle database

```
SELECT *
FROM ARCMBTransaction
WHERE transaction_date >= TO_DATE('2023-07-21', 'YYYY-MM-DD')
  AND transaction_date < TO_DATE('2023-07-22', 'YYYY-MM-DD')
   OR transaction_date >= TO_DATE('2023-08-17', 'YYYY-MM-DD')
  AND transaction_date < TO_DATE('2023-08-18', 'YYYY-MM-DD')
   OR transaction_date >= TO_DATE('2023-09-12', 'YYYY-MM-DD')
  AND transaction_date < TO_DATE('2023-09-13', 'YYYY-MM-DD')
   OR transaction_date >= TO_DATE('2023-10-22', 'YYYY-MM-DD')
  AND transaction_date < TO_DATE('2023-10-23', 'YYYY-MM-DD')
   OR transaction_date >= TO_DATE('2023-11-15', 'YYYY-MM-DD')
  AND transaction_date < TO_DATE('2023-11-16', 'YYYY-MM-DD')
   OR (transaction_date >= TO_DATE('2023-12-23', 'YYYY-MM-DD') AND transactio
n_date <= TO_DATE('2023-12-31', 'YYYY-MM-DD'));
```

Annex 2 Mobile money transaction data capture from the back end database

```
import pandas as pd
from google.colab import drive
# Mount your Google Drive
drive.mount('/content/drive')

# Specify the path to your CSV file on Google Drive
file_path = '/content/drive/My Drive/MBTransaction.csv'

# Read the CSV file into a DataFrame
df = pd.read_csv(file_path)
# Display the first few rows of the DataFrame
print("First few rows:")
print(df.head())
# Display the last few rows of the DataFrame
print("\nLast few rows:")
print(df.tail())
print("\n\n\n\n\ the uploading ends here ")
```

Annex 3 Mobile banking data loading

```
import pandas as pd
from google.colab import drive
# Mount your Google Drive
drive.mount('/content/drive')

# Specify the path to your CSV file on Google Drive
file_path = '/content/drive/My Drive/MMTransaction.csv'

# Read the CSV file into a DataFrame
df = pd.read_csv(file_path)
# Display the first few rows of the DataFrame
print("First few rows:")
print(df.head())
# Display the last few rows of the DataFrame
print("\nLast few rows:")
print(df.tail())
print("\n\n\n\n\ the uploading ends here ")
```

Annex 4 Mobile money data uploading

```
Mounted at /content/drive
First few rows:
    MOBILE_APP_ID CHANNEL_NAME        CUSTOMER_ID     ACCOUNT_ID  \
0     ussdpayment       Mobile  C-FF5C7AA53D70D2A  1000177644831
1     ussdpayment       Mobile  C-FF5C3EB3E594B6E  1000160441619
2     ussdpayment       Mobile  C-FF5C8ADC99C70A1  1000371753897
3  androidpayment       Mobile  C-FF5CC8F8AC0A430  1000450647218
4  androidpayment       Mobile  C-FF5CCE89908CEC2  1000401117869

           ACCOUNT HOILDER    PROCESSED_DATE  \
0       HALMA SULEMAN NURU  30/01/2024 21:49
1          EFTU AHMED MUME  30/01/2024 14:50
2      MAHLET GIRMA DEMISE  30/01/2024 14:50
3     ABEL TEFERI MEKONNEN  30/01/2024 14:50
4  YIDIDIA TADESSE MEKURIA  30/01/2024 14:50

                                        PRODUCT_NAME       RECIPIENT_ID  \
0  ETB 6,000.00 debited from HALMA SULEMAN NURU f...  C-FF5C7AA53D70D2A
1  ETB 1,000.00 debited from EFTU AHMED MUME for ...  C-FF5C3EB3E594B6E
2  ETB 560.00 debited from MAHLET GIRMA DEMISE fo...  C-FF5C8ADC99C70A1
3  ETB 700.00 debited from ABEL TEFERI MEKONNEN f...  C-FF5CC8F8AC0A430
4  ETB 300.00 debited from YIDIDIA TADESSE MEKURI...  C-FF5CCE89908CEC2

                         VENDOR_ID     PRODUCT_ID           RECIPIENT_NAME
\
0  FF5AC8C6CE55EC1-1-1000177644831  1000250151555                 MB05HSN
1  FF5AC8C6CE55EC1-1-1000160441619  1000368100149               eftumb789
2  FF5AC8C6CE55EC1-1-1000371753897  1000539032554                 MAH551
3  FF5AC8C6CE55EC1-1-1000450647218  1000435552971    ABEL TEFERI MEKONNEN
4  FF5AC8C6CE55EC1-1-1000401117869  1000377713177  YIDIDIA TADESSE MEKURIA

  STATUS_CODE TRANSACTION_TYPE                              SELECTED_OPTIONS
0        PAID            Debit     param.decimal.Amount: 6000; Reason: HA;
1        PAID            Debit      param.decimal.Amount: 1000; Reason: 1;
2        PAID            Debit     param.decimal.Amount: 560; Reason: 560;
3        PAID            Debit  param.decimal.Amount: 700; Reason: getish;
4        PAID            Debit     param.decimal.Amount: 300; Reason: tty;

Last few rows:
        MOBILE_APP_ID CHANNEL_NAME        CUSTOMER_ID     ACCOUNT_ID  \
1048570   ussdpayment       Mobile  C-F6B27FB433FF9E3  1000112951761
1048571   ussdpayment       Mobile  C-FF5C652A754CC02  1000375409074
1048572   ussdpayment       Mobile  C-FF5CB05A0EEA862  1000275549472
1048573   ussdpayment       Mobile  C-FF5C867F32F9BC3  1000394280902
1048574   ussdpayment       Mobile  C-FF5CCF0B8F4D4CE  1000273943394

                ACCOUNT HOILDER    PROCESSED_DATE  \
1048570  GETENET WONDEMU WOLEDE  30/01/2024 16:54
1048571       BELAYNESH G/HIWOT  30/01/2024 16:54
1048572     SARA ACHISO DINDAMO  30/01/2024 16:54
1048573          NURU SEID ABGAZ  30/01/2024 16:55
1048574  MELIKEYAS BEZABHE MENA  30/01/2024 16:56

                                             PRODUCT_NAME
RECIPIENT_ID  \
1048570  ETB 456.8 debited from GETENET WONDEMU WOLEDE-...  C-
F6B27FB433FF9E3
1048571  ETB 5.00 debited from BELAYNESH G/HIWOT-ETB-90...  C-
FF5C652A754CC02
1048572  ETB 50.00 debited from SARA ACHISO DINDAMO-ETB...  C-
FF5CB05A0EEA862
```

Annex 5 Result of data loading and display mobile banking

```
# Access the 'SELECTED_OPTIONS' column and display it
selected_options = df['SELECTED_OPTIONS']
print(selected_options)
```

```
0                    param.decimal.Amount: 6000; Reason: HA;
1                     param.decimal.Amount: 1000; Reason: 1;
2                    param.decimal.Amount: 560; Reason: 560;
3                 param.decimal.Amount: 700; Reason: getish;
4                    param.decimal.Amount: 300; Reason: tty;
                                  ...
1048570               param.decimal.Amount: 450; Remark: 1;
1048571    param.decimal.Recharged Mob No: 0952546283; pa...
1048572    param.decimal.Recharged Mob No: 0929627185; pa...
1048573              param.decimal.Amount: 20000; Reason: 1;
1048574    param.decimal.Recharged Mob No: 0916605607; pa...
Name: SELECTED_OPTIONS, Length: 1048575, dtype: object
```

➔End➔