2020

Detecting financial crimes using pattern recognition techniques: case of mobile money transactions.

Eshiwani, Michelle Mercy
Faculty of Information Technology
Strathmore University

Detecting Financial Crimes using Pattern Recognition Techniques: Case of Mobile Money Transactions

By

Michelle Mercy Eshiwani

112819

A Thesis Submitted to the Faculty of Information Technology in partial fulfillment of the requirements for the award of Master of Science in Information Technology.

Master of Science in Information Technology

Strathmore University

March 2020

Declaration and Approval

I Michelle Mercy Eshiwani declare that this research has not been submitted to any other University for the award of a Degree in Master of Science in Information Technology. This Thesis does not contain any content that was produced by another person except where due reference is made in the Thesis itself.

Student Name: Michelle Mercy Eshiwani

Sign: _____ Date: _____

Supervisor's Name: Dr. Vincent Omwenga

Sign: _____ Date: _____

Abstract

Financial Crimes have evolved and gained complexity in the recent past owing to advanced technological adoption globally. As consumers have accepted new forms of service delivery that offer them convenience, affordability and easy access, criminals have also found new avenues of pushing their illegal funds or financing criminal activities without raising suspicion or being detected.It is therefore widely recognised that the prevalence of economically motivated crime in many societies is a fundamental threat to the development of world economies and their stability.

This research aimed to develop a pattern recognition tool to analyze transaction patterns and detect suspicious transactions.This would in turn reduce the impact of financial crimes on mobile money transactions in terms of loss of revenue for both individuals, corporations and countries by safeguarding legitimate transactions while also tying any loose ends that facilitate the transfer of illegally acquired funds over legitimate channels.This research focused on the field of Pattern Recognition in identifying and analyzing fraud in mobile money transactions. The tool applied Statistical Pattern recognition using the K-Nearest Neighbor algorithm to accurately classify transactions as fraudulent or genuine.

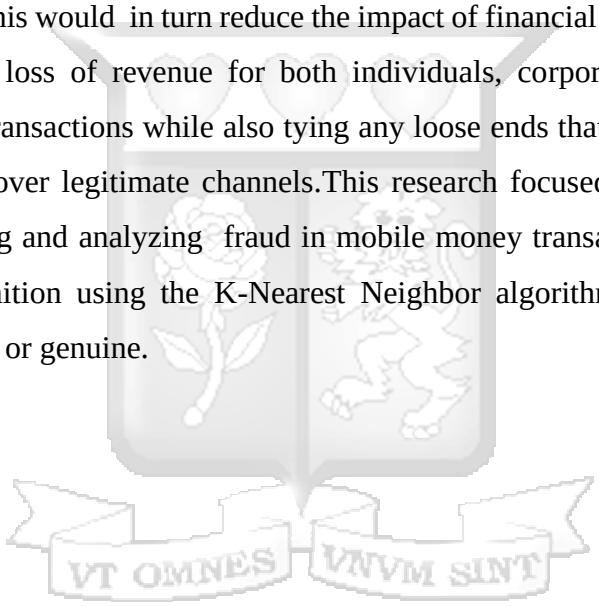


Table of Contents

Abbreviations and Acronyms

AML	-	Anti-Money Laundering
AML/CFT	-	Anti-Money Laundering/ Combating the Financing of Terrorism
CDD	-	Customer Due Diligence
CFT	-	Combating the Financing of Terrorism
FIU	-	Financial Intelligence Units
KYC	-	Know Your Customer
ML	-	Money Laundering
M-MONEY	-	Mobile Money
MMT	-	Mobile Money Transactions
MNO	-	Mobile Network Operator
SIM	-	Subscriber Identification Module
TF	-	Terrorist Financing

Definition of Terms

Financial Crimes - Financial crimes are categorized as financial abuse crimes thet are non-violent in nature but result in the loss of an individual or entity's financial security (Jung & Lee, 2017). It encompases, among others Fraud, Money Laundering (ML) and Terrorist Financing (TF).

Pattern Recognition -The science concerned with the classification of data into different categories based on similarities in already existing knowledge (Bishop, 2006).

SIM- Subscriber identity module; refers to the smart card used in mobile phones. It carries the user's identity for accessing the network and receiving calls and stores personal information, such as phone directory and short-message service messages received (Theodorou & Okong'o, 2019).

Synthetic Data- Any production data not obtained by direct measurement and is considered anonymized. Created by stripping any personal information (names, license plates, etc.) from a real dataset so it is completely anonymized. Conceptually, synthetic data may seem like a compilation of "made up" data, but there are specific algorithms designed to create realistic data. Synthetic data can assist in teaching a system how to react to certain situations or criteria (Lopez-Rojas, 2016).

PaySim Synthetic Mobile Money Transaction dataset- A dataset generated using the PaySim simulator that uses aggregated data from a private dataset to generate a synthetic dataset that resembles the normal operation of transactions and injects malicious behaviour to later evaluate the performance of suspicious activity pattern recognition methods (Lopez-Rojas, 2016)

List of Figures

List of Tables

.

Chapter 1: Introduction

1.1 Background

Financial crimes have been recognized as a global challenge with impactful economic and social ramifications.Jung & Lee, (2017) categorized financial crimes as financial abuse, that is non-violent in nature but results in the loss of an individual or entity's financial security,with the potential to jeopardize global economies and homeland security. These crimes include, among others fraud, Money Laundering (ML) and Terrorist Financing (TF).

These crimes thrive on relevant volumes of financial transactions to conceal the identity, source, and destination of illegally gained money. Experts estimate that up to $2 trillion of illicit proceeds from human trafficking, bribery and fraud flow through legitimate financial system (Didimo & Liotta, 2014).

Authorities are therefore constantly looking out for new ways and means to track down and prevent these crimes, just as much as criminals are developing innovative tactics in order to stay ahead. In a bid to tackle financial crimes, most governments have at the local level established special cross collaborative investigative agencies known as (FIUs).They conduct money laundering (ML), terrorism financing (TF) and asset tracing investigations (UNCTAD, 2018).

As one of the newest forms of financial services, Mobile Money, initially enabled people to make basic cash transactions on their phones. However, increased popularity of mobile money services has fuelled major growth in sectors of the economy, including financial institutions, retail and wholesale traders, agriculture, education and health.Most service providers have integrated mobile money platforms into their payment systems owing to their convenience and speed (Rolfe, 2019).

In Kenya , a global trend setter for mobile money services ,registered mobile money customers are able to receive services such as mobile credit, sports betting, insurance, cross-border remittances, bill and utility payments, airtime top-ups, and savings from the comfort of their mobile wallets. The Central Bank of Kenya (CBK) reported that the sum cash transferred using mobile money and payment services topped KES2.87 trillion ($27.7 billion) in the period to end-August of 2019, up 10.3 per cent on the same period of 2018.This means that Kenyans moved nearly half the equivalent of the country's gross domestic product (GDP) through their mobile

phones in the last financial year, underlining the growing importance of digital wallets to the economy (Pham, 2019).

While trying to effetively cater to customers evolving needs, mobile money has unfortunately also become a conduit for financial crimes. Owing to high transaction speeds, reduced physical contact with agents (except for cash deposits or withdrawals) and layering multiple transactions in small margins,Mobile Money Providers are often playing catchup to secure transactions as well as their customers from the complex risks they pose (UNCTAD, 2018).

Mobile Network Operators (MNOs), the providers of mobile money platform services such as M-pesa, Airtel Money, T-Kash and Equitel, are mandated by the Central Bank of Kenya and the Communications Authority (CA) who review the systems and regulations in place to safeguard transactions and customers and ensure that they meet the "highest global standards". MNOs are subject to the Proceeds of Crime and Money Laundering Act, 2012 and are therefore mandated to regularly file reports on suspicious with the country's Financial Reporting Centre. Besides monitoring all their transactions to detect and report suspicious activity, MNOs state that their partners (in the case of both local and international remittances) are equally required to undertake due diligence on remitters in line with their respective countries' fraud and anti-money laundering regulations (Masinde , 2017).

According to Masinde (2017), Monitoring and investigating suspicious transactions within mobile money transactions remain difficult despite some checks already in place. This is because criminals have utilized the fact that mobile money is less regulated in comparison to traditional financial institutions to carry out their business.Unlike banks where one cannot transact beyond a certain threshold without the approval of the central bank, mobile money customers can transact large sums of money using multiple SIM cards and at various agents undetected.This brings into question the integrity and sustainability of mobile money transactions particularly as the entire value of funds moved through these platforms rises.

A Failure detect and deter fraudulent transactions reduces any apparent consumer advantage gained as well as financial inclusion growth in these markets. Furthermore, regulatory authorities as a result may be less disposed to allow the necessary space to expand and diversify innovations on mobile financial services if they view MNOs internal controls as insufficient in detecting and mitigating financial crimes (Buku & Mazer, 2017).

1.2 Problem Statement

According to KPMG (2019), financial crimes are often concealed in complex patterns, to hide the motive, source of funding and even the identity of the criminal. As is the case of fraudulent transactions on mobile money platforms, they often go undetected, because their characteristics are not a 'one size fits all' as is mandated by the Anti-Fraud ,AML and CFT guidelines that that financial institutions must adhere to.

Standard rule-based systems currently in place to monitor transactions most often only identify these crimes after they have occurred, and the perpetrators have been able to clean up any traces of evidence. This is because the method used by the criminals do not meet the predefined system threshold that would have them flagged.

Applying pattern recognition to mobile money transactions for continuous monitoring was necessary especially in identifying patterns in customers transactional activities. To act as a benchmark of expected patterns of legitimate transactional behaviour, used to accurately detect unusual or potentially suspicious transactions.

1.3 Aim

The purpose of this research was to develop a tool for detecting financial crimes in mobile money transactions using pattern recognition to differentiate between genuine and suspicious transaction.

1.4 Specific Objectives

(i) To investigate the characteristics of mobile money financial transactional activities.
(ii) To review the effectiveness of current financial crimes detection and controls systems.
(iii) To analyse Pattern recognition as a tool for detecting financial crimes in Mobile Money Transactions.
(iv) To design and develop a tool for detecting financial crimes based on pattern recognition in mobile money transactions.
(v) To test the ability of the tool to detect a financial crime on a mobile money platform.

1.5 Research Questions

(i) What are the characteristics of mobile money transactional activities?
(ii) How effective are current financial crimes controls and systems?

(iii)How is Pattern recognition applied to financial crime detection?

(iv)What pattern recognition techniques can be applied in crime detection for mobile money transactions?

(v) How effective are pattern recognition techniques in detecting financial crimes on mobile money transactions?

1.6 Justification

As the total value of funds transferred through these mobile money platforms continue to increase, there is significant seriousness in the need for early detection and investigation of financial crimes. Adequate preventative measures must be in place to ensure the integrity and sustainability of these transactions, which are currently the biggest conduit of financial crimes in Kenya.

The presence of such a system would ensure continued consumer benefit and continued financial inclusion gains. Furthermore, allowing regulators to encourage more innovations to expand and diversify mobile money services.

1.7 Scope and Limitation

This research was entirely focused on identifying the nature of financial crimes in mobile money transactions. The study looked into statistical pattern recognition tools to facilitate the analysis and detection of fraud in mobile money transactions in Kenya.

Chapter 2: Literature Review

2.1 Introduction

This chapter focuses on the relevant literature that aids in understanding the research problem: detecting suspicious transactions through pattern analysis of large volumes of data in mobile money transactions. It further presentes the various empirical studies and theoretical frameworks of technologies applied or proposed in detecting suspicious transactions in various financial transactions.

2.2 Mobile Money Transactions

Subia and Martinez (2014) described Mobile Money as an ecosystem of various types of financial activities or services transacted on a mobile device.The service facilitates the transfer of cash to the digital wallet, while encouraging the adoption of many more innovative services built on its foundations. In Kenya, Mobile money is a service provided by Mobile Network Operators (MNO) such as Safaricom (M-pesa), Airtel Kenya (Airtel Money) and Telkom Kenya (T- Kash) and regulated by the Communications Authority of Kenya.

GSMA (2017) acknowledges that since being adopted, especially in developing countries, Mobile Money has greatly contributed to economic growth and financial inclusion of people and communities who were "unbanked".Individuals from the informal sector have been provided with easily accessible and affordable financial services that were previously the preserve of those in the formal working sector .

Mobile money services are categorized according to the type of transactions and how they should be processed. They are currently categorized as.

(i) **Mobile payments-** A service that enables registered vendors to receive payments for the purchase of goods and services from customers using their mobile wallet (on condition that available balance is enough) via mobile device. This service extends to utility payments, retail payments, government collection and payments etc. (Bettcher & Mihaylova, 2015).

(ii) **Mobile transfer** -A service that allows registered customers to send or receive money to or from any customer registered on the same network or across networks on condition that interoperability is enabled. For a transfer to take place ,a customer must either deposit

cash into their mobile wallet through a registered Mobile Network agent or via USSD from bank (Ombaka, 2018).

(iii)**Mobile banking**- Mobile banking allows customers of financial institutions (banks ,Saccos, insurance firms) to transact form their accounts to mobile money wallets and vice versa via USSD or integrated applications.Customers have access to a wide array of financial services deposits ,withdrawals ,utility payment (Bettcher & Mihaylova, 2015) .

2.2.1 Characteristics of Mobile Money Transactions

According to Lal & Sachdev (2015), mobile money services globally have many similarities in their approach to service delivery. Highlighted in the subsections below are the distinct characteristics that define mobile money transactions.

2.2.1.1 Instant transactions

Transactions are processed within seconds, as opposed to within hours or business days as is the case with traditional FIs. Instant speeds allow payments flexibility, making funds available, while increasing the control of personal and business funds (Valchev, 2019).

2.2.1.2 Seamless Integrations

Integration between merchants, institutions and MNOs has been improved to ensure seamless flow of transactions to business wallets, banks and institutions (Valchev, 2019).

2.2.1.3 Transactions

Except for cash deposits or withdrawals which are dependent on the presence of agents, all other transactions do not require any third-party intervention.

2.2.1.4 Security

Mobile wallets are secured by a range of robust technologies, such as point-to-point encryption, tokenization, passwords, biometrics, out-of-band authentication, one-time password (OTP) via SMS, security questions.In Kenya customers must input a unique four pin code to successfully complete a transaction.

2.2.1.5 Transaction Charges

These are operational charges that are applied to user accounts based on the amount of cash being transacted. In Kenya transaction charges are set at the discretion of MNOs with the guidance of the Communications Authority (CA) and are changed successively depending of the amount range being transacted (Bahia & Muthiora, 2019).

2.2.1.6 Transaction Limits

Transaction limits refer to the threshold amount that Mobile Money customers can transact either per given transaction, daily, weekly and monthly. These limits are set by the regulating authority for AML/CFT tracking (Bahia & Muthiora, 2019).

2.2.2 Financial Crimes targeting Mobile Money Transactions

Anti-money laundering, fraud and Combating the financing of Terrorism regulations/guidelines in place for Traditional Financial Institutions may not legally apply to the new industry entrants that facilitate m-money because their core business is communication services. The m-money market is generally newer than financial crimes legislation in many countries, and governments did not consider m-money and its unique operations when drafting these laws (Chatain, *et al*. 2011).

Poor oversight on a regulators part intensifys anonymity, elusiveness, rapidity (risks posed by mobile money transactions). According to Chastain, *et al.* (2011), further complicating the problem is determining the right government authority to oversee m-money. In Kenya, Mobile Money is regulated by the Communications Authority and not the Central bank of Kenya.

In Kenya, mobile money is regulated by the Communications Authority of Kenya (CA). However, there is a push by Members of Parliament to delink mobile money services from their parent telecommunication firms and be registered as separate commercial banks. If the law makers have their way, the telecommunications regulator, (CA), would be compelled to ensure that mobile money services like Safaricom's M-Pesa, Airtel Money and Telkom's T-Kash are licensed as banks hence come under the jurisdiction of the Central bank of Kenya and its regulation (Mutai, 2019) .

A sample of potential vulnerabilities at each stage of M-money transactions for the different risk category is provided in Table 2.1 below.

Table 2.1: Possible vulnerabilities for different risk types for M-Money transactions (Solin & Zerzan, 2010)

General risk factor	**Example of vulnerabilities for different transactions**		
	Deposit	**Transfer**	**Withdrawal**
Anonymity	A criminal can open Multiple accounts, with falsified identification documents and using different identities to hide the true nature of deposits. Mobile money agents are not adequately equipped to verify the authenticity of a customer's identification documentation during sim card registration.	Suspicious names are not recognised by the system, making it a safe zone for known criminals and terrorists.	withdrawal of illegitimate or terrorist-linked funds is possible especially where a mobile money agent shows laxity in in verifying a customer's identification during the transaction. Mobile money agents are not adequately equipped to verify the authenticity of a customer's identification documentation during sim card registration.
Elusiveness	Criminals can openly redirect illicit funds into multiple accounts.	They can carry out multiple transactions to confuse the money trail and the true origin of funds.	Redirected funds from multiple accounts can be withdrawn at the simultaneously.
Rapidity	Illicit funds can be quickly as is a characteristic of mobile money transferred to different accounts.	Transactions occur in instantaneously, hence difficult to flag and screen for suspicion of terrorist financing or money laundering.	Just as in deposits, withdrawals are also done fast from different accounts.
Poor oversight	Without proper guideline and regulation, mobile money services pose a great systemic risk.		

2.2.3 Systems and Controls used to secure Mobile Money Transactions

This section discussed the systems and controls systems and controls that were in place to secure mobile money platforms and agents and customers.

2.2.3.1 System and Controls that Secure the Mobile Money Platform

Fraud Management System

MNOs have implemented fraud management systems to detect fraudulent transactions based on geolocation technologies that flagged transactions based on crime hotspots such as prisons (Field, 2012).

Access Security

Generally, MNOs require customer authentication for a transaction to be executed. Access security can be in the form of inputting a preset pin, OTP-based authorizations (from third party applications), Mobile Station International Subscriber Directory Number (MSISDN) for international remittances and Public Key Infrastructure (PKI) for authenticating users and devices during online transactions (Mahindra Comviva, 2016).

2.2.3.2 System and Controls that secure Customer and Agent Transactions

This section described the basic security controls commonly available on the customer and agent side.

Establishing User Identification

When a user registers to be a Mobile money customer, they are issued with a PUK (Personal unlocking key) which is unique to every user. The PUK is used to reset user PIN (Personal Identification Number) when they forget it. A customer also gets a four digit numerical unique PIN which can be changed the PIN from the M-PESA Tool Kit for a charge (Mule, 2015).

Confirmation Messages

Once a customer performs a withdrawal from an agent, a confirmation message is sent to both the Agent and customer to verify the transaction details after which cash is released. For a deposit, a customer first hands over the cash to the agent who then goes ahead to deposit the cash to the customers mobile wallet both parties also receive similar confirmation messages (Mule, 2015).

Prompt Response

Once money is withdrawn, a customer receives a prompt whether they would like withdraw money from that agent. This prevents customers from withdrawing money from a wrong agent.

When withdrawing money, a customer must input the correct agent number to ensure the transaction is successful (Mule, 2015).

Back rolling Transactions

A customer has the ability to roll back a transaction if deposited into a wrong account.The reversal process is done within seconds after the transaction .

2.3 Characteristics of Financial Crimes

Croall (2005) agreed that financial crimes are the objects or the target of illegal and often prohibited means to obtain the personal benefit from the illegal conversion of the ownership of the property of others.The are charaterised by type of crime, victim and perpetrator as elaborated in the subsections below.

2.3.1 Type of crime

2.3.1.1 Fraud

Defined as false representation for a criminal's personal gain, where fraudsters doorstep tactics to target their victims via communication media such a phone, email, or communication sites.

Consumer affecting frauds: Consumers fall victim to frauds such as identity theft, social engineering scams, rogue agents, loss from incorrect transfer to inadvertent beneficiaries who are unwilling to relinquish the cash (Buku & Mazer, 2017).

Agent affecting frauds: Agents also fall victim to frauds such as float loss in the agent's account arising from unauthorized use, compromised PINs and ploys involving impersonation of MNO staff to gain unlawful access to the agent's float account. Customers can also perform withdrawal reversal fraud or deposit fake currency (Buku & Mazer, 2017) .

Internal Fraud in MNOs: Internal fraud has caused substantial losses for MFS providers, while putting at risk user accounts while raising integrity concerns for the system. For example, in Kenya, Safaricom employees facilitated fraud on the platform leading to loss of clients' money running into millions of shillings (Kiplagat, 2020).

According to Buku and Mazer (2017) insufficient internal controls and audit processes, poor corporate constructs, a lack of employee awareness on fraud, and inadequate whistle blowing

systems are among the significant contributors to internal fraud.

2.3.1.2 Money Laundering

Money Laundering (ML) as defined by Anosh & Ahmadi (2015) ,is the method by which criminals try to legitimize their profits from criminal activities (drug trafficking, people trafficking, embezzlement, corruption etc) into the legitimate financial world. The FATF produced guidelines for financial institutions on how such suspicious transactions should be handled.

2.3.1.3Terrorist Financing

Terrorist financing involves the solicitation, gathering or delivery of resources with the intention that they may be used to support terrorist actions or groups. More precisely, according to the International Convention for the Suppression of the Financing of Terrorism, a person commits the crime of financing of terrorism "if the person by any means, directly or indirectly, unlawfully and wilfully, affords or gathers funds with the intent that they should be used, in full or in part, in order to carry out" an offense within the scope of the Convention.The primary goal of individuals or entities involved in the financing of terrorism is therefore not necessarily to conceal the sources of the money but to conceal both the financing and the nature of the financed activity (IMF, 2012).

Omondi (2019) wrote on court proceedings in Kenya against suspected masterminds of a terrorist attack at dusitD2 hotel in Nairobi that claimed 21 lives have exposed vulnerabilities in mobile money transfer services. Two Suspected terrorists had registered several mobile phone numbers that were used to receive Ksh. 109 million collectively to finance terrorist activities. The cash would then be withdrawn through an M-Pesa till numbers at the Diamond Trust Bank Eastleigh Branch.The money would then be funneled to the terrorist group 'Al Shabaab' in Somalia who claimed responsibility for the attack. One of the accused, a registered Mpesa agent registered a total of 52 fake accounts in a span of two months to aid in distributing these funds.

A third defendant in the case , the bank Manager was faulted for her failure to flag the large volumes of transactions as suspicious, given that there are regulations which require a Mobile Payment Service Provider or its agent to set transaction or payment account limits. According to regulations, any account exceeding a daily turnover of KSh100,000 and any personal account

transacting more than KSh300,000 per week should be investigated.The Agent and the MNO however did not flag these accounts (Omondi, 2019).

2.3.2 Categories of victim

This section looked at the nature of the criminal or victim that makes prosecution and punishment more or less likely.

2.3.2.1 Members of the Public

Reporting and successive investigation of financial crimes such as fraud are subject to the number of victims involved and the quantity of money in question. The greater the number of individuals affected or the money lost the more likely the crime is to be investigated .Also taking into consideration the education background, age and economic standing of the victim (Croall, 2005).

2.3.2.2 Mobile Network Operator

If the fraud is perpetrated internally, and the MNO is the victim, investigations on the suspicious transactions will be done internally to maintain customer confidence. According to Buku and Mazer (2017) insufficient internal controls and audit processes, poor corporate constructs, a lack of employee awareness on fraud, and inadequate whistle blowing systems are among the significant contributors to internal fraud.

2.3.3 Category of Criminals

This section discussed the categories of individuals or groups of individuals who perpetrate and profit from financial crimes.

2.3.3.1 Politically Exposed Persons (PEP)

A Politically exposed Person is rated as an individual with a high risk of bribery, corruption, and money laundering by virtue of their position within in the state. This term was coined by FAFT who recommend that enhanced monitoring of accounts should be implemented for such individuals (Croall, 2005).

2.3.3.2 Rogue Insiders: Major and Petty

Insider frauds suffered by MNOs such as employees colluding with criminals to defraud customers by illegally sharing out their transactional data. Such cases often question the ability

of MNOs to secure customers transactions as well as attract non-compliance fines of not putting in place adequate measure to cater to internal and external crimes (Croall, 2005).

2.3.3.3 Organised crime groups

This section refers to a network of a criminals who stage well organised attacks on m-money customers and agents. Using scams such as social engineering, these criminals take advantage of the naivety of their victims to defraud them of money in their mobile wallets (Buku & Mazer, 2017).

2.4 Systems and Controls used to detect Financial Crimes in Traditional Financial Institutions

Traditional financial institutions have the advantage of experience in terms of the systems and controls they implement to secure transactions and monetary compliance. The following are the systems and controls currently in place in place for detecting financial crimes.

2.4.1Transaction Monitoring

Transaction monitoring systems facilitate Financial Institutions in monitor customer transactions for AML/CFT risk. They combine and analyze this information together with a customers' account profile, to determine a customer's profile, risk levels, and predicted future activity. The transactions monitored can include cash deposits, withdrawals,payments and transfers, (Comply Advantage, 2018).

2.4.2 Know Your Customer (KYC) and Customer Due Diligence (CDD) Systems

Know Your Customer (KYC) and Customer Due Diligence (CDD) is an area of regulatory requirement. These systems and controls allow FIs to focus their compliance efforts on determining their customers risk index. Figure 2.1 below elaborates the processes of KYC and CDD.

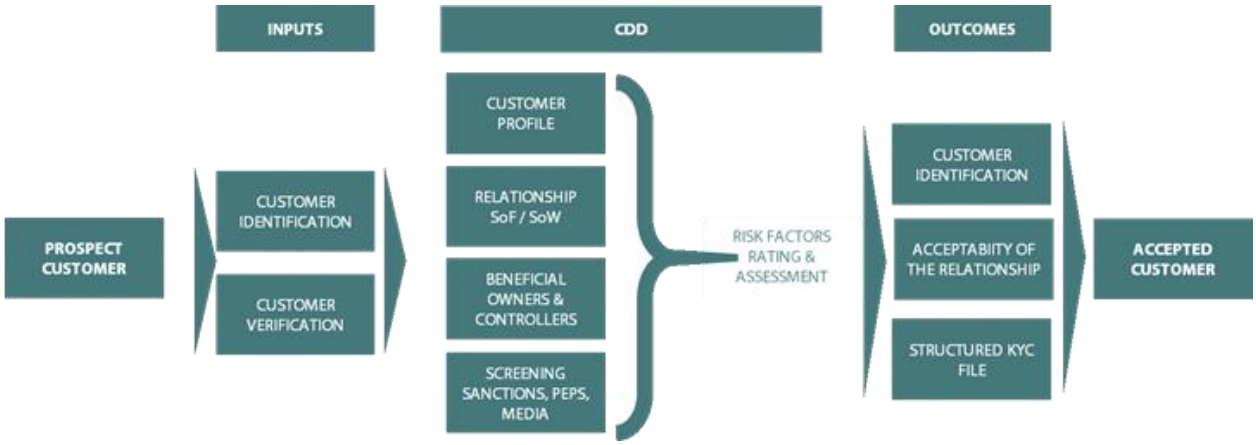


Figure 2.1: Customer Due Diligence Process (Fisher, 2017)

During customer onboarding, KYC and CDD are focused on the authentication and substantiation of customer identity to establish their risk rating. Continuous due diligence requires persistent monitoring of customer transactions to detect suspicious activities Enhanced KYC systems should ideally integrate from end-to-end the customer process framework covering on-boarding requirements, continuous risk monitoring and reporting (Chartis Research, 2015).

2.4.3 Sanctions and watch-list monitoring

Sanctions screening is a control employed within (FIs) to detect, prevent and manage sanctions risk as per regulatory compliance. Screening is undertaken as a Financial Crime Compliance (FCC) programme, to assist in identification of sanctioned individuals and organisations and the illegal activity to which a financial organization can be exposed.Reference databases for Money Laundering, corruption and terrorism are used to rate customers based on police reports ,news articles etc. as linked to them. (Wolfsberg Group, 2019).

The Wolfsberg Group (2019) advise that FIs must continuously update watchlists and ensure low false (positive or negative) reporting rates while updating the information and applying it in real time. An excessively cautious system can lead to high false positive reporting while systems that only screen for direct matches have a high false negative reporting. Figure 2.2 displays the risk approach used to screen customers.

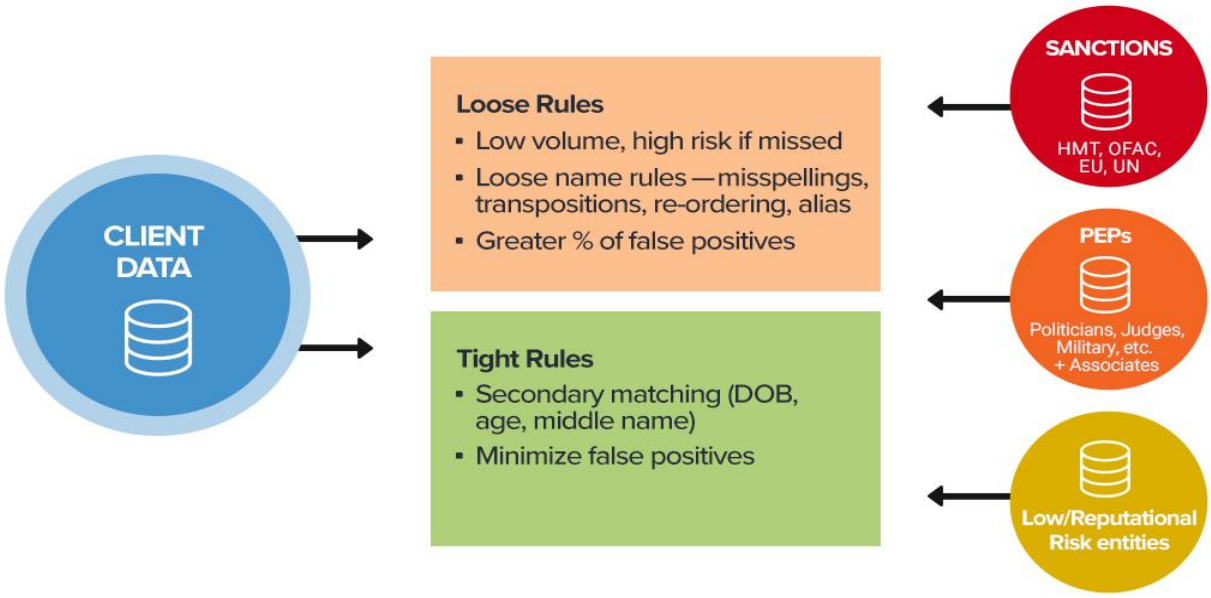


Figure 2.2: Risk based approach in sanction screening (FinScan, 2016)

Financial institutions are fined for sanctions violations,therefore forcing them to be constantly up to date with all sanctions, fraud or trade monitoring guidelines, in any jurisdiction through which they carry out business (Ernst and Young, 2016).

2.4.4 Anti-Money Laundering (AML) and Combating the Financing of Terrorism (CFT) Compliance

Money laundering is the process by which the illicit source of income generated by criminal activities is concealed to hide the link between the funds and the original crime. On the other hand, Terrorist financing involves raising and processing assets to supply terrorists with funding to facilitate their criminal activities (Owuor, 2016).

Anosh and Ahmadi (2015) discussed that while these two crimes differ in many ways, they often exploit the same vulnerabilities in financial systems that allow for an inappropriate level of anonymity and non-transparency in the execution of financial transactions.

Due to the sensitivity of financial custodianship, all traditional Financial Institutions in Kenya are mandated by by the Central bank of Kenya to comply with FATF (Financial Action Taskforce) Recommendations in handling their AML and CTF processes. FATF is an internationally endorsed global standard for implementing effective AML/CFT measures to facilitate transparency, traceability, and accountability within the banking industry

(FATF*GAFI, 2010).

(i) The activities undertaken by the taskforce are:

(ii) Setting international standards to combat money laundering and terrorist financing.

(iii) Assess and monitor compliance with the FATF standards.

(iv) Conduct typologies studies of money laundering and terrorist financing methods, trends and techniques.

(v) Responds to new and emerging threats, such as proliferation financing.

These activities increase the transparency of a financial system while providing members with the capacity to successfully counter money launderers and terrorist financiers (Financial Action TAsk Force FATF*GAFI, 2010).

2.4.5 Cyber-Security

Financial institutions are frequently stepping up their efforts in responding to occurrences that are often directed at multiple channels, products and systems to improve their cyber security controls (Buku & Mazer, 2017).

2.5 Effectiveness of Controls and Systems in Detecting Financial Crimes in Kenya

Financial crimes vary depending on the industry that an organisation operates in, their jurisdictional risk, the products and services they offer and how mature their compliance operations are. Controls within financial services have enforced a semblance of decorum amongst most customers in terms of compliance and customer sensitization . However, there is a constant need to evaluate and evolve to cater to the complexity and rapidity of financial crime evolution (Financier WorldWide, 2018).

Many firms have underinvested, or not made their AML/CFT programs a firm priority. Therefore, allowing criminals to thrive on their illegitimate profits. It is not enough to implement emerging and innovative technology such as transaction monitoring tools, also improving the quality of data and leveraging public sources of data for validation, are all critical strategies in successfully detecting suspicious transactions.

2.6 Pattern Recognition

Pattern Recognition is the science or process concerned with the classification of data into different categories based on similarities in already existing knowledge. Its ultimate goal is to optimally extract patterns based on certain conditions and to separate one class from the others (Bishop, 2006).

2.6.1 Pattern Recognition Process in Suspicious Transaction Detection

The pattern recognition process is composed of preprocessing, feature extraction, and classification as illustrated in figure 2.3. A data source or dataset is preprocessed, so that it becomes suitable for subsequent sub-processes. The next step is feature extraction, in which, the dataset is converted into a set of feature vectors which are supposed to be representative of the original data. These features are used in the classification step to separate the data points into different classes based on the problem (Karyakarte1 & Savant, 2019).

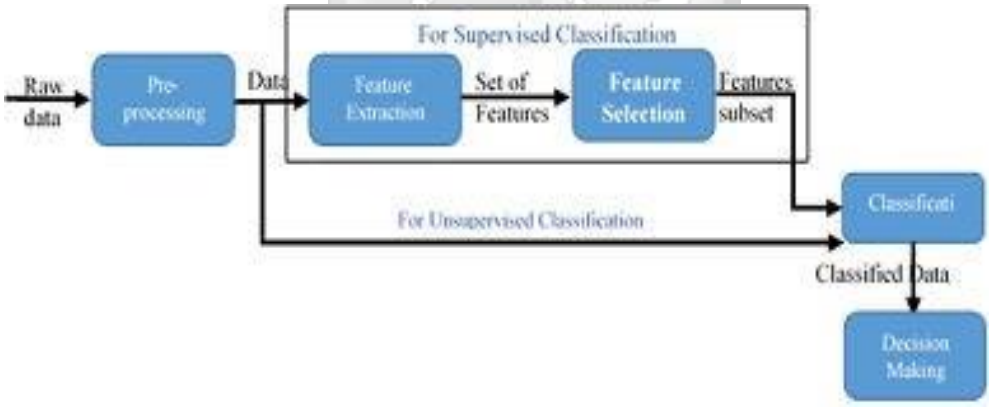


Figure 2.3: Pattern Recognition Process (Liu, SUn, & Wang, 2006)

2.6.1.1 Pre-processing

Tang (2014) defines the role of preprocessing as segmenting the unique pattern from the background. It is used to reduce variations and produce a more consistent set of data. Preprocessing should include some noise filtering .

2.6.1.2 Feature extraction

A feature is the measure of observable data corresponding to a pattern.Feature extraction is used to overcome the problem of high dimensionality of the input set in pattern recognition. As in figure 2.3 above, the input data is transformed into a reduced representation set of features, also termed as feature vector. Only the relevant information from the input data should be extracted in order to perform the desired task using this reduced representation instead of the full size input (Karyakarte1 & Savant, 2019).

Features extracted must be easily computed, robust, rotationally invariant, and insensitive to various distortions. Then optimal features subset that can achieve the highest accuracy results should be selected from the input space.

2.6.1.3 Feature selection

According to Guyon and Elisseeeff (2003),the purpose of feature selection is categorised into three processes:

(i) Refining the detection operation.

(ii) Delivering faster and more affordable detectors.

(iii) Providing a better understanding of the underlying process that data was generated from.

The features extracted at extraction phase are put through a filtering process to acquire a more discriminative feature vector. At this point the physical meaning of the original features is maintained. The feature vector available at the end of this step is the training data .

2.6.1.4 Classification

According to Sharma and Kaur (2013), classification is the process of of assigning a label to an input item with the help of an algorithm. The input items are the feature vectors produced after feature selection.

As illustrated in figure 2.4, if a classification algorithm acknowledges a refined feature set from the feature selection step as input, then it is a supervised classification algorithm.The classifiers that contain the knowledge of each pattern category and also the criterion or metric to discriminate among patterns classes.However in the case of an unsupervised classification

algorithm,if the system parameters are adapted using only the information of the input, and constrained by prespecified internal rules. It attempts to find inherent patterns in the data that can then be used to determine the correct output value for new data instances (Karyakarte1 & Savant, 2019).

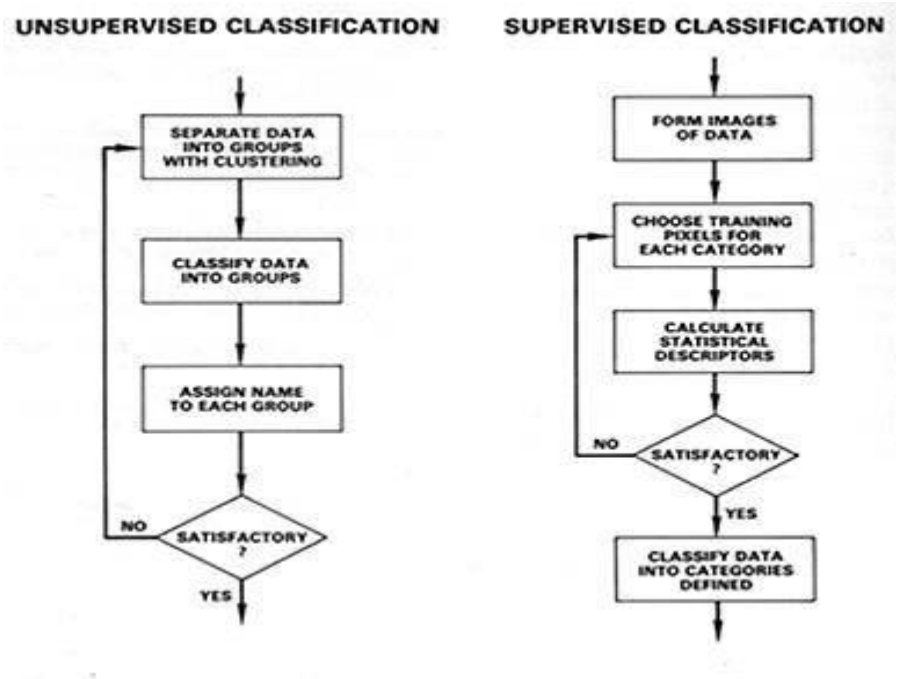


Figure 2.4: Unsupervised and Supervised Classification (WGBIS, 2014)

2.6.1.5 Decision making process

The input of this stage is the classified data (supervised or unsupervised classification) ,and provides system users with reliable results to make decisions on how to proceed with flagged transactions (Sharma & Kaur, 2013).

2.6.2 Review of Pattern Recognition Techniques used to detect Financial Crimes.

Pattern recognition techniques are divided in to three categories i.e., Supervised pattern recognition techniques, semi-supervised pattern recognition techniques and unsupervised pattern recognition techniques. (Asht & Dass, 2012).

2.6.2.1 Supervised Techniques

Supervised pattern recognition uses supervised learning algorithms to create classifiers from various object classes (Gujral *et al.*, 2019).

Statistical Pattern recognition

Statistical pattern recognition systems are extensively because of their simplicity, i.e. they are based on statistics and probabilities. Traits are recoded in form of numbers which are then applied to create a pattern. Each pattern is therefore represented by a specific multidimensional vector, used for pattern recognition (Duin et al., 2002).

The methods/algorithms that are applicable for statistical pattern recognition include; Naïve-Bayes (NB-C), Linear Discriminant Analysis (LDA-C), Kernel Classifier (Kernel-C), Least Mean Square Linear Classifier (Linear LMS-C), Least Mean Square Quadratic classifier(PolQuadraticLMS-C), Multinomial logistic regression model with a ridge estimator(Logistic-C) and Particle Swarm Optimization - Linear Discriminant Analysis (PSOLDA-C).

Challenges with Statistical Pattern Recognition

Issues with representation (similarities), feature reduction and classifier complexity(adaptation) and classifier training and imbalanced data (generalization) can easily impact on the precision of the pattern recognition model (Aguilar, 2004).

2.6.2.2 Unsupervised Techniques

Unsupervised classification finds hidden features in unlabeled data using clustering or data segmentation techniques (Gujral *et al.*, 2019).

Template matching Pattern Recognition

Template matching is generally used in image processing .This describes how patterns are identified by clusters of pixels or curves to localize and identify shapes in image. Thus, patterns are identified in form of templates (Raj *et al.*, 2015).

Challenges with Template matching Pattern Recognition

This technique is only suitable for matching image patterns and not transactions.

2.6.2.3 Semi-Supervised techniques

Semi-supervised techniques define a more complex relationship between elements.

Structural /Syntactic Pattern Recognition

Structural or syntactic pattern recognition focuses on the identification of a hidden pattern and is achieved by matching its symbolic representation with several predefined object models. In the structural approach, the association is made by a characteristic match that calculates a measure of similarity between the unknown input and several prototype models. In syntactic pattern recognition, a parser or error-correcting parser checks an unknown input for its accordance with the rules of a grammar that describes all members of a pattern class (Serratosa et al., 2011).

Challenges with Syntactic Pattern Recognition

The syntax of the language is not known explicitly, only a sample of patterns is given hence

defining such an algorithm is much more difficult. Then again, a lack of a grammatical inference algorithm makes the use of a syntactic pattern recognition model impossible in most of the real-world applications (Jain et al., 2000).

2.7 Empirical Literature on Pattern Recognition in Detecting Financial Crimes

This section covers some of the most relevant works related to the application of pattern recognition techniques in the field of detecting crimes.

There is limited research in the use of pattern recognition in detecting financial crimes specifically in mobile money transactions. AML/CFT tools developed for financial institutions mostly use data mining techniques for suspicious transaction detection. To the best of the researcher's knowledge, there are no applications involving real-time detection of financial crimes on mobile money transactions. Nevertheless, there are works that have applied pattern recognition to flag fraud in the financial industry. Unfortunately, these articles do not specify the databases used.

A pattern in crime detection is defined as a series of crimes committed by the same offender or group of offenders. To identify true patterns, one would need to consider information beyond simply time and space, but also other features of the crimes.There are few known previous works aimed directly at detecting specific patterns of financial crime in mobile money transactions (Wang & Rudin, 2013).

2.7.1 A Conceptual Framework for Detecting Financial Crime in Mobile Money Transactions

Gombiro, Jantjies and Mavetera (2015), in their proposed framework were influenced by the fact that existing methods of financial crime detection did not address the occurrence of false positive rates adequately. They proposed to overcome this limitation using big data analytics, where the method looked at large volume of data and dealt with a variety and removal of noisy data. Based on the millions of transactions generated by M-money transactions, there exists a need to extract value from the data by differentiating legal from illegal transactions on mobile money transactions.

2.7.1.1 Proposed Framework

The framework encompassed monitoring activities such as fraud, social network analysis and money laundering. Built on the foundation FATF 2012 recommendations and preliminary observations using M-money platforms and their requirements.

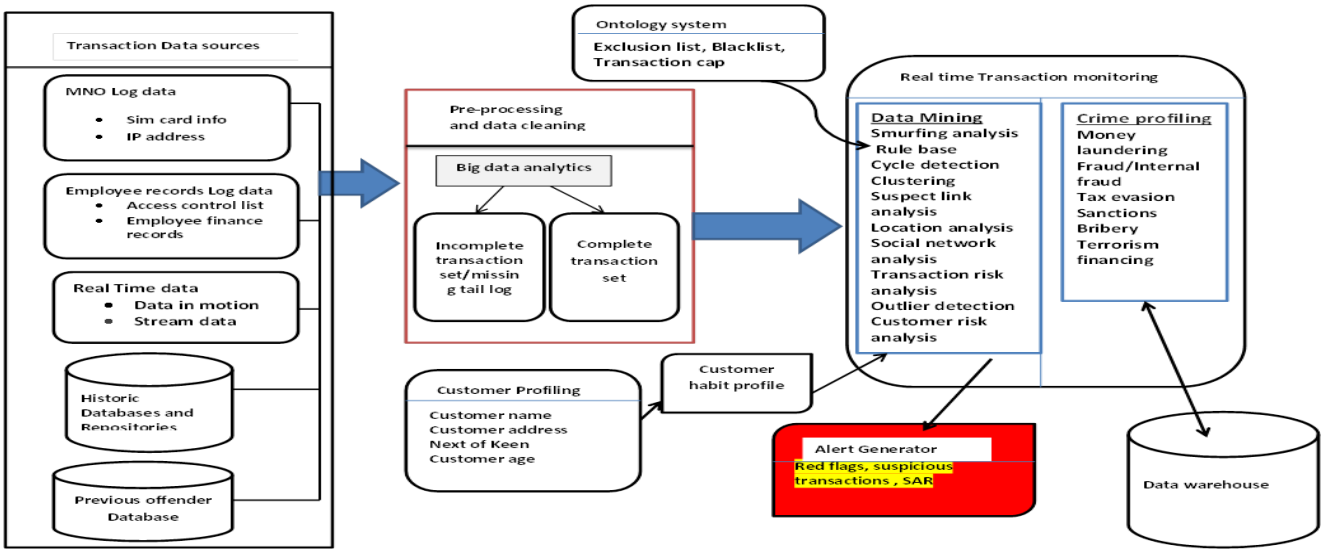


Figure 2.5: Conceptual for detection of M-money financial crime

The framework applied data cleaning, pre-processing and use of historical databases. Input data was sourced from employee logs, historical offenders' database, and historical and real time transaction data. The flags and Suspicious Activity Reports (SAR) were then forwarded for analysis to a crime analysis agency for further analysis. Ontology was set to define a set of rules to link customers, customer ranking, threshold values and transaction rules (Gombiro, Jantjies, & Mavetera, 2015).

2.7.1.2. Success and Drawback

The framework addressed the problem of identifying anomalies in mobile money transactions by use of various data mining approaches as well as the use of Dempster Shafer theory to identify evidence of financial abuse and assign probabilities based on likelihood of transaction to fall under illegal transaction.

However, the researchers acknowledged that they could have selected a different algorithm for faster execution in the identification of suspicious transactions and execution of the tools. Furthermore, there is a need to continuously update the rule base as criminals always come with new techniques in money laundering, cybercrimes, and other financial crimes.

2.7.2 Predicting Fraud in Mobile Money Transfer

Adedoyin (2018) in his research, proposed a pattern recognition model to predict fraud in Mobile money transfer transactions using Case-Based Reasoning (CBR).The researcher deduced that most machine learning techniques depend on statistically relevant datasets for prediction. However, in the absence of a significant size of historical data, they would most often not perform well .

Hence the application of Case-based reasoning (CBR) , a modern computational method that solves new problems using solutions (specific knowledge) from past and similar problems that were successfully solved .

2.7.2.1 Proposed Framework

Figure 2.6 below illustrates the architecture of the proposed framework which the researcher

implemented to detect mobile money transfer fraud. The case-based reasoning (CBR) system was proposed as the classifier due to the absence of historical consumption data.

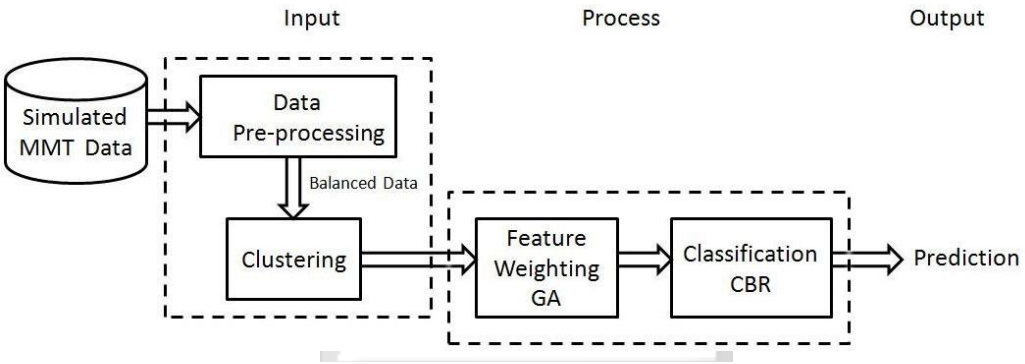


Figure 2.6: Fraud Detection Framework

The proposed framework consisted of three main components: Input, Process and Output detailed as follows.

Input: Under the framework, the simulated MMT data were first pre- processed by running several queries to verify it's quality .A clustering algorithm was then applied to guide and reduce the search space by collecting the most similar clusters. This enabled cases collected under similar circumstances to be identified and limited the retrieval to these cases.

Process: The CBR classifier was used to classify new instances of MMT transactions as either fraud or non-fraud cases. To exploit the flexibility of weighting all the input vectors in the process component, a Genetic Algorithm (GA) was used to calculate their weight to reflect the significance of each vector as determined by the GA.

Output: As indicated in Figure 5.1, the output section provided a summary window for the prediction which displayed the ranking of clusters of transaction neighbors.

2.7.2.2 Success and Drawback

From the results of the experiments, it was determined that the pattern recognition model could not only identify suspect transactions, as well as provide the ranking of clusters of transaction neighbors for new cases. CBR with clustering approach showed a better time performance but with significantly low accuracy, which was the drawback of the research.

2.7.3 Fraud Detection in Mobile Money Transactions Using Machine Transactions Using Machine Learning

Kang (2019) in this study explored a data mining system for fraud detection in mobile financial transactions. The researcher compared two supervised machine learning models, random forest and gradient boosting, for their applicability in the detection of fraudulent records. The primary data source was the Paysim Simulator for detecting fraud in mobile money transactions.

2.7.3.1 Proposed Framework

The research aimed to build a probing system, by adopting supervised learning, where known normal and fraudulent cases were used to train the models to learn their characteristics.

The researcher partitioned the raw data frame into two subsets, training and testing sets. The former trained random forest and gradient boosting with labeled data, making the systems exploit the patterns of legal and illegal transactions.The systems were able to predict which class a new observation belonged to. The models were then applied to the testing set, to verify both methods while evaluating their accuracies.

2.7.3.2 Success and Drawbacks

The accuracies of the models were higher than expected, due to the synthetic nature of the dataset. In the real-world environment, the accurately would be lower. The researcher should have tested a sample of real-world transaction to truly verify the accuracy of the output.

2.8 Summary of Empirical Literature on Detecting Pattern Recognition in Detecting Financial Crimes

Table 2.2 is a summarization of studies both significant and small that influenced the researcher in developing the pattern recognition tool for detecting financial crimes on mobile money transactions.

Table 2.2 Summary of Empirical literature on Detecting Financial Crimes using Pattern Recognition.

Study	**Model or framework**	**Drawbacks**
A Conceptual Framework for Detecting Financial Crime in Mobile Money Transactions by Gombiro, Jantjies and Mavetera (2015).	Big data analytics	The framework required an algorithm that executed faster in the identification of suspicious transactions. Furthermore, there is a need to continuously update the rule base as criminals always come with new techniques in financial crimes.
A Framework for Predicting Fraud in Mobile Money Transfer by Adedoyin,A (2018).	Case-Based Reasoning (CBR) with clustering.	CBR with clustering approach showed a better time performance but with significantly low accuracy
Fraud Detection in Mobile Money Transactions Using Machine Transactions Using Machine Learning by Kang (2019)	two supervised machine learning models, random forest, and gradient boosting	Data used was highly imbalanced therefore bringing into question the true accuracy of the model's ability to detect fraudulent transactions.

Autoregressive-based outlier algorithm to detect money laundering activities by Kannan and Somasundram (2017).	Autoregressive-based outlier algorithm.	The inability to differentiate between normal and suspicious transactions was a limitation. Also, the use of a Linear Regression algorithm resulted in lack of deep analysis of the problem and the time consumption.
A Multi-Variant Relational Model for Money Laundering Detection using the Time Series Data Architecture by MCA & Prakaran (2014)	Relational mapping by differentiatin multiple accounts and a time series by splitting the data at a particular time frame	A criminal can however deposit illegitimate funds into different accounts to legitimize it.

2.9 Gap Analysis

Traditional financial crime detection systems and controls are primarily dependent on the use of data mining for focused identification, verification and customer profiling .For Mobile money transactions, this is however not the case as KYC and CDD is only done at customer onboarding at which point an agent cannot really authenticate the customers identity.

To gain strategic advantage in this fight against financial crimes, adopting solutions that do not merely respond to past patterns of attack, but in addition are highly anomaly aware,in analyzing and detecting suspicious patterns in real-time.

2.10 Conceptual Framework

The continuing threat of financial crimes in mobile money transactions reinforces the necessity for deploying pattern detection and an alerting functionality on the actual traffic flowing through. The framework encompasses monitoring activities such as fraud, terrorism financing and money laundering.

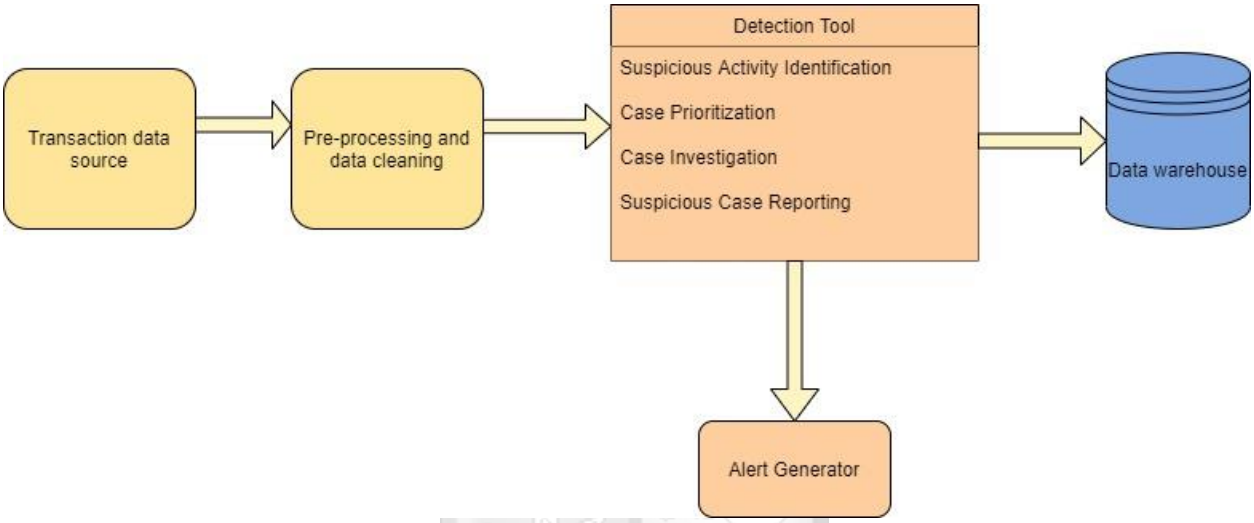


Figure 2.5: Conceptual Framework

The data from the data source underwent pre-processing to remove any noisy transactions while ensuring that only complete transactions were forwarded for monitoring.Feature extraction was subsequently performed from the initial set of measured data and was completed with the necessary features to facilitating the successive learning and generalization steps. Realtime transaction monitoring continuously applied activity monitoring for suspicious transactions. The typical four stages in monitoring were:Suspicious activity identification after which the tool would send alerts to system users upon the detection of a suspicious transaction. Finally, the data would then be stored for future reference and link analysis.

Chapter 3: Research Methodology

3.1 Introduction

This section described the main research methodology that was adopted in carrying out this research. Structured System Development (SSD) is a formal process of eliciting system requirements, both to reduce the possibility of the requirements being misunderstood and to ensure that all the requirements are known before the system is developed. It also introduces rigorous techniques to the analysis and design process (Wells , 2009).

3.2 Research Design

The research design adopted in this study was exploratory-descriptive due the fact that it was important to identify the commonalities in data (explorative) and to categorize the commonalties in a particular manner (descriptive) (Brink & Marilynn , 1998).

The combination of these two elements was used to extract insights out of the data. The analysis of the common factors and their correlation uncovered details in the subject matter that was critical in understanding it.

3.2.3 System Development

The system development applied for this research was Agile Methodology. This methodology allows for repeated improvements on the different modules of the system based on the success of the research and the discovery of new technologies to improve the functionality of the anticipated tool. Above all, it enabled the researcher to better define the system requirements as the process was done incrementally (Lu & DeClue, 2011).

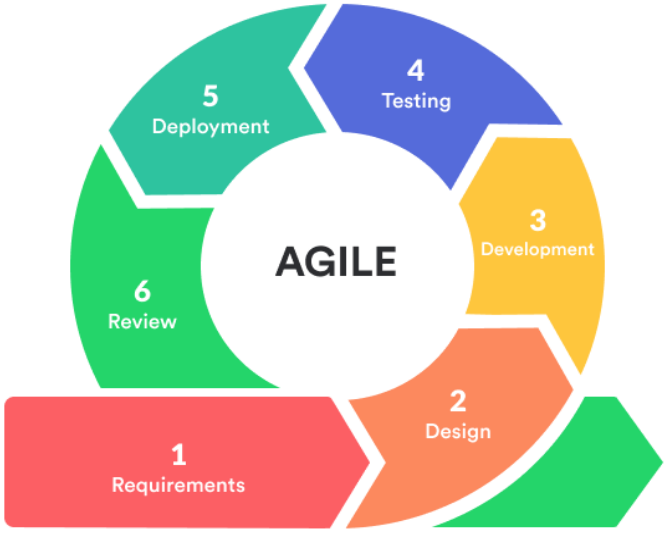


Figure 3.1: Agile System development

The Agile development lifecycle involves the basic steps mentioned below.

3.2.3.1 Requirements

This phase involved capturing requirements for the iteration based on thorough document analysis. These requirements were

(i) Realtime detection of suspicious transactions
(ii) The tool should be able to detect suspicious transactions such as fraud, money laundering and terrorist funding.
(iii) Accurate reporting of suspicious transactions

3.2.3.2 Design

The design stage involved the use of high-level UML diagrams and wireframes o demonstrate how the tool would function and how it would fit into the already existing system.

3.2.3.3 Development/Iteration

Once the requirements were defined, developing iterations of the project began, with the goal of having a working product to launch at the end of the sprint.

The tool was developed using Python as a programming language due to its vast number of available libraries. Matlab Pattern Recognition Toolbox for representation and generalization SQLlite database was used to store data on suspicious transactions.

The product underwent various rounds of revisions; therefore, this first iteration includes only the bare minimum functionality. There is an allowance for future additional sprints to expand upon the overall system.

3.2.3.4 Testing

Testing involved validation of the accuracy of the model for the final detection tool using and the second phase of testing was conducted on the tool using test data reserved from the dataset. Confusion matrix was also applied.

3.2.3.5 Deployment

This phase involved releasing the live version of the system. The production phase ends when support has ended or when the release is planned for retirement.

3.2.3.6 Review

This stage involved reviewing the results of the first round of testing the model and reworking them into the requirements of the next iteration.

3.2.4 System Analysis

This phase involved the studying the existing (Transaction monitoring system) and the proposed framework in context of their interrelationship and eliminate redundancies.

3.2.5 System Design

System design is expresses the architecture, components, modules, interfaces and data for a system to satisfy specified requirements (Waldo, 2006).

This research made use of UML diagrams such as context diagram, Data flow diagrams (level 1&2), Data Model, Database schema and finally wireframes (to depict the GUI of the tool). These aids were a suitable representation of the proposed tool regarding SSD Methodology.

3.3 Target population and Sampling

Asiamah, Mensah and Oteng-Abayie (2017) defined target population as a group of individuals or participants with specific attributes of interest and relevance to a research study. The target population of this research was financial transactions conducted on mobile money platforms. Due to data privacy laws and the difficulty in obtaining such data, a synthetic dataset, Paysim Mobile money simulator was used.

The target population required a data source with both numerical and categorical features like transaction type, amount transferred, account numbers of sender and recipient accounts etc. The Paysim Synthetic dataset for mobile money transactions sourced from the Kaggle website, is the only data source that contained such detailed transactional data. The entire dataset frame consisting of 6,362,620 observations, 11 columns and five transaction types, was sampled during data analysis.

3.4 Data collection

The primary source of transactional data of this research was a synthetic mobile money dataset, the Paysim Mobile money simulator. The data set simulates mobile money transactions based on a sample of real transactions extracted from one month of financial logs from a mobile money service implemented in an African country. The original logs were provided by a multinational company, a Mobile Network Operator currently running in more than 14 countries around the world.

3.5 Data Pre-processing

Data directly mined from the Paysim Mobile money simulator contained some missing and erroneous data, therefore making immediate analysis impossible. Several transactions contained with zero balances in the destination account both before and after a non-zero amounts were transacted. Such data underwent transformations and were then recorded for data analysis

3.6 Data Analysis

Exploratory Data Analysis (EDA) was most appropriate in determining how to improve and come up with better threshold to capture the suspicious transactions. The outcome of the data analysis was that the target variable 'isFraud', which was the actual fraud status of the transaction while 'isFlaggedFraud' was the indicator which the simulation used to flag transaction using some threshold.

3.7 Research Quality

This tool was evaluated and validated to ensure that the results are reproducible and stable.

3.7.1 Reliability

Application of the Confusion matrix, Classification report and the Area Under Receiver Operating Characteristic (AUROC) indicated that the model applied in the tool was able to detect substantially more true positive fraudulent transactions than false positive.

3.7.2 Validity

Validity was concerned with the degree to which the research findings were applied to the real world unbiasedly, beyond the controlled setting of the research. The tool was able to accurately classify fraudulent transactions from a random selection of dummy data.

3.7.3 Ethical Considerations

This research was based on the original ideas of the researcher and any externally borrowed concept included in the research was fully referenced and cited in-text to acknowledge the source of the data and their contribution to the research.

In addition, the data collected was secondary in nature and was sourced from the Paysim Mobile money simulator dataset, which does not avail any unique identifiers that would link an actual individual to a transaction. Any actual personal information had already been redacted from the dataset hence no chance of any data privacy violation.

Furthermore, the research was issued an Ethical approval by the Strathmore university Institutional Ethical Review committee, as well as a Research Permit from the National Commission for Science and Technology and Innovation (NACOSTI) after due consideration.

Chapter 4: System Analysis and Design

4.1 Introduction

This chapter expounds on the analysis and design of the Pattern recognition tool detecting financial crimes on mobile money platforms, by incorporating the various requirements that were identified after successful data collection and analysis.

It further defines the different stakeholders of the system, system components, system data models and system process models. This chapter also captures the visual representation of the proposed solution using visual modelling language (UML) to ensure the tool was completely understood before development.

4.2 Data Analysis

The goal of this analysis was to better understand the synthetic mobile money transaction dataset, pre-process it and create a model for prediction. The data analysis method applied was exploratory data analysis (EDA) as earlier stated in Chapter 3. The Paysim mobile money Dataset for fraud detection, comprised of 6362620 observations and 11 columns, table 4.1 is a summary of the variables/labels in the dataset.

Table 4.1 Summary of Variables (Kang, 2019)

	Variable Name	Format Example	Description
1	step	5	Each step is an hour of time in real world. The largest number for step is 744 (the 30th day)
2	type	PAYMENT (Categorical variable)	Transaction types (CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER)

3	amount	8424.74	Transaction amount in local currency
4	nameOrig	C1000001725	Customer who started the transaction
5	oldbalanceOrig	351422.72	The initial balance of sender before the transaction
6	newbalanceOrig	257557.59	The new balance of sender after the transaction
7	nameDest	M1974356374	Customer/Merchant who received the transaction
8	oldbalanceDest	526950.37	The initial balance of receiver before the transaction
9	newbalanceDest	771436.84	The new balance of receiver after the transaction
10	isFraud	1 (Categorical variable)	The status of a transaction (0 as legitimate and 1 as fraudulent)
11	isFlaggedFraud	0 (Categorical variable)	The status that the system identified for a transaction — here an attempt to transfer more than 200,000 (in local currency) in a single transaction will be flagged as an illegal attempt (0 as normal and 1 as illegal attempt)

4.2.1 Transaction Count

The Transaction count graph provides a visualisation of the count for each type of transaction as listed below.

CASH_OUT -2237500, PAYMENT-2151495, CASH_IN-1399284, TRANSFER-532909, DEBIT -41432 (Name: type, dtype: int64)

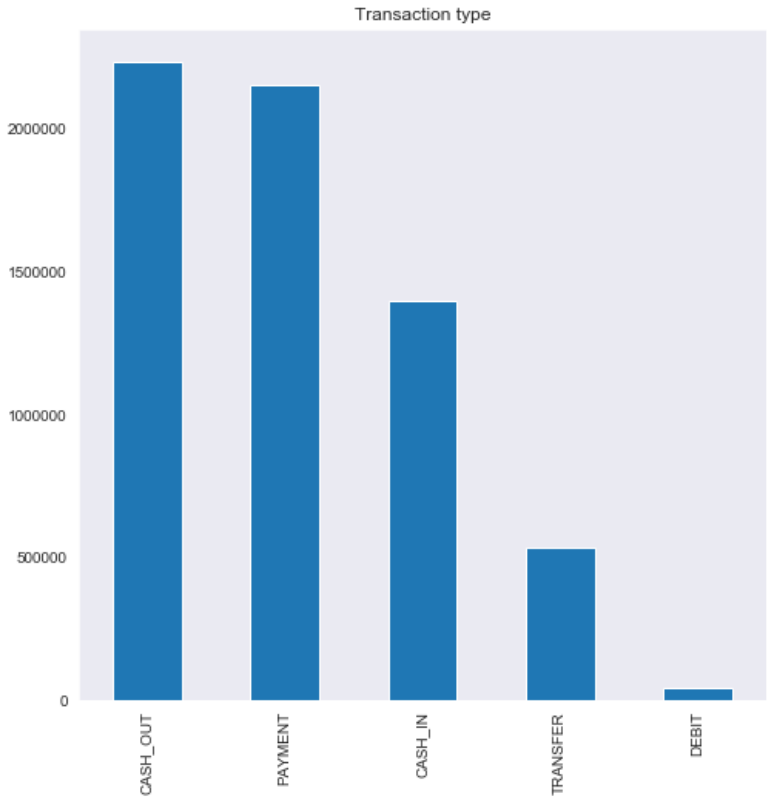


Figure 4.1: Transaction count per type

4.2.2 Transactions types flagged as suspicious

Figure 4.2 below indicates the actual number of each transaction type that were flagged as fraud. The snippet in figure 4.2 displays the transactions flagged as fraudulent.

```
The types of suspicious transactions are ['TRANSFER', 'CASH_OUT']

The number of suspicious TRANSFERs = 564

The number of suspicious CASH_OUTs = 578
```

Figure 4.2: Transaction types flagged as suspicious

4.2.3 Handling Imbalanced Data

Unrelated transaction types were filtered out and only what was relevant was maintained. Fraud only existed in 0.3% of the dataset, indicating that the data was highly imbalanced. To handle this problem, data under-sampling method was used to prevent bias towards the majority class under sample the dataset.

```
print("The fraud transaction of the filtered dataset: {0:.4f}%".format((len(tmp[tmp.isFraud == 1])/len(tmp)) * 100))

The fraud transaction of the filtered dataset: 0.2965%
```

Figure 4.3: Handling Imbalanced Data

4.3 Requirements Analysis

This research aimed at developing a pattern recognition tool for financial crime detection. Based on this objective, the succeeding sections outline the various requirements for the proposed solution. The requirements were mainly gathered through document analysis by the researcher.

4.3.1 Functional Requirements

These are the key system functionalities of the pattern recognition tool that must meet user specifications by identifying the tasks and activities that must be accomplished. They include:

(i) There should be an integration between the proposed tool and existing Mobile Money System to pull transactional data real time.

(ii) The tool should detect fraudulent transactions based on analysis of a customers of transacting patterns.

(iii) The tool should enable update database with flagged transactions for future link analysis.

(iv) The tool should be able to send alerts in real time when a transaction is flagged, via email to authorised system users.

4.3.2 Non-Functional Requirements

These describe the constraints under which the tool must work within, hence the following considerations.

(i) Usability

The system is intended for use by the internal compliance and audit team of any MNO firm. to improve their financial reporting capability and hence improve regulatory compliance, service delivery and platform integrity.

(ii) Data Security

The data being processed by the system is confidential and should be treated as such hence the need for user rights and roles on data management to be assigned as per a company's data policies.

(iii)Persistent Storage

Configuration to the SQLlite database server was key for the researcher, to ensure uninterrupted connectivity when historical records are being accessed.

4.4 System Process

The proposed system process in Figure 4.6 illustrates the general layout of the Pattern recognition tool. The major steps that take place in the system are as follows:

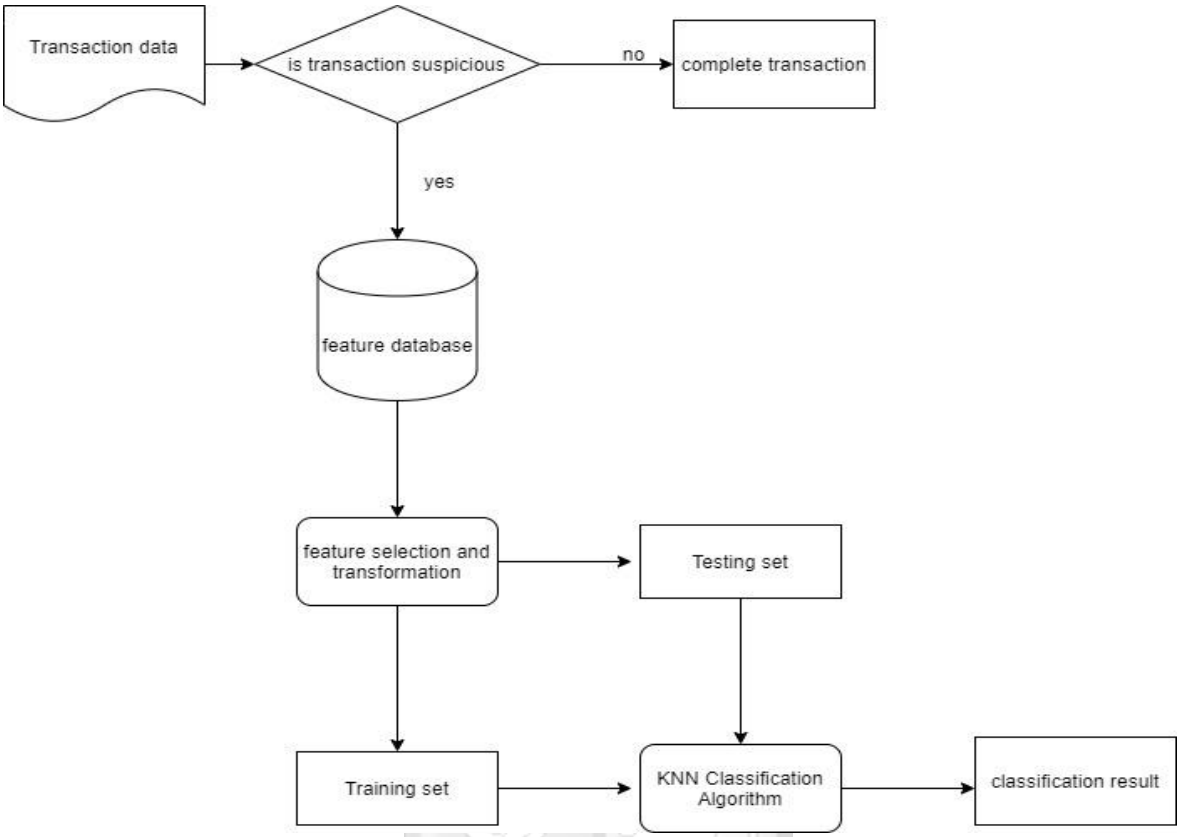


Figure 4.4: System Architecture

4.5 Data Flow Diagrams

DFDs represent the flow of data of the proposed tools process and provide information about the outputs and inputs of each entity and the process itself.

4.5.1 Context Diagram

The context diagram in figure 4.7 depicted the boundary of the tool, its environment and the entities that it interacted with i.e. Mobile Money system and system users. Additionally, it showed the inputs (Transaction data source) and outputs (analysed results) from the various entities.

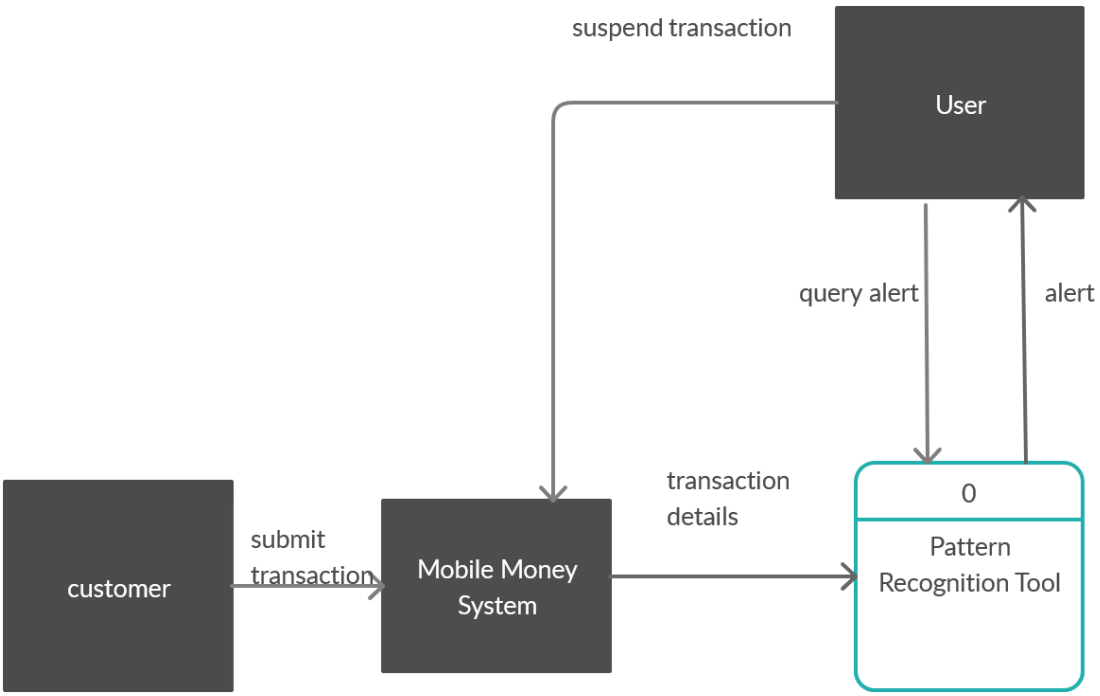


Figure 4.5: Context Diagram

4.5.2 Data Flow Diagram Level 1

The level 1 DFD notated each of the main sub-processes that together form the complete system. An authorized system user upon login would be able to query the system for transactions that were flagged as fraudulent within a given date range.

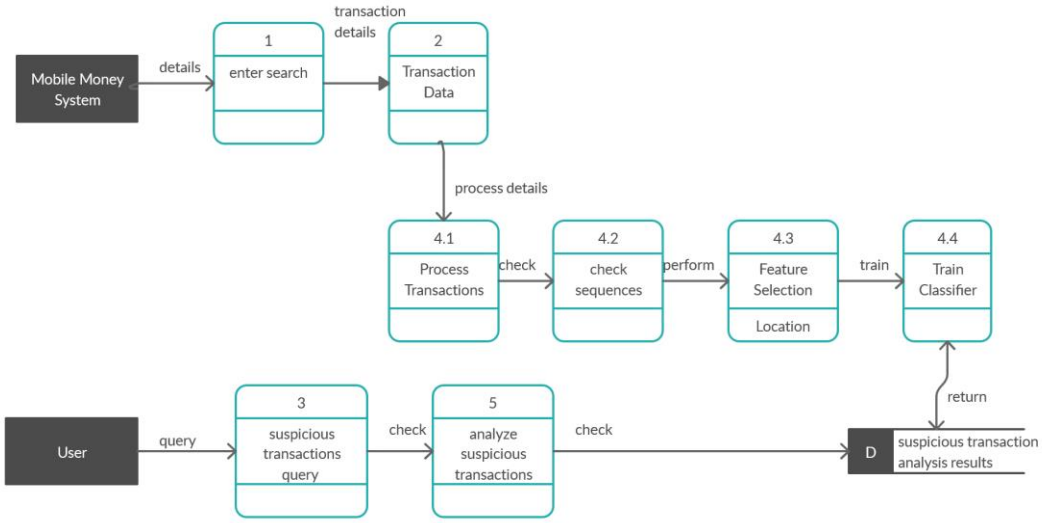


Figure 4.6: Data Flow Diagram Level 1

4.5.3 Data Flow Diagram Level 2

The level 2 data flow diagram (DFD) offered a more detailed look at the processes that made up the system in comparison to level 1 DFD does.

Once a transaction is flagged as fraudulent, a system user can carry out an analysis of the customer's transactional history and update their risk rating (based on the various sanctions lists) as part of the customer due diligence process.

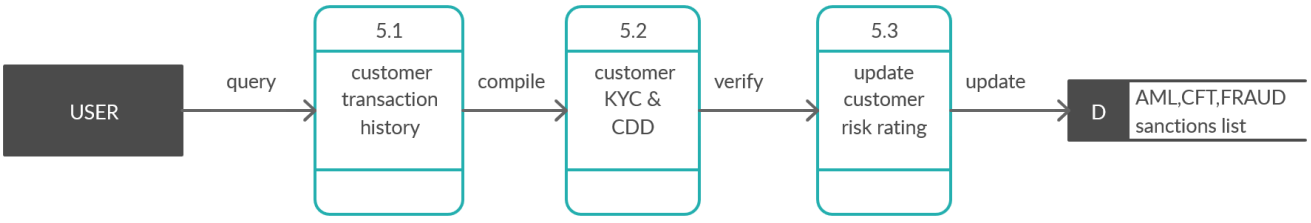


Figure 4.7: Data Flow Diagram Level 2

4.6 Data Model

The data model was an abstract model that organized the elements of data exactly how they related to one another and to the properties of other real-world entities.

The database was made up of the customer, transaction, transaction type and system user tables, each with its induvial attributes and how they related to each other as in figure 4.8.

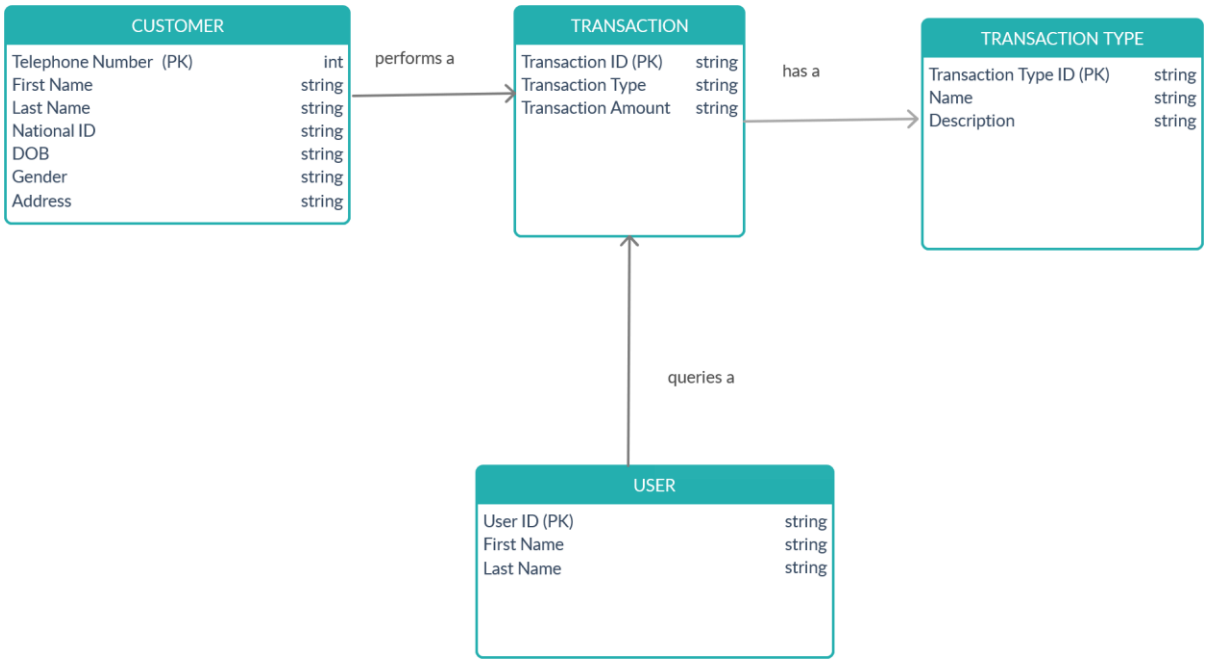


Figure 4.8: Data Model

4.7 Database Schema

The database schema defined how the data was organized and how the relations among the tables were associated.

The data was organized into four entities, customer, transaction, transaction type and system user as well as how each entity associated with the other captured in figure 4.9.

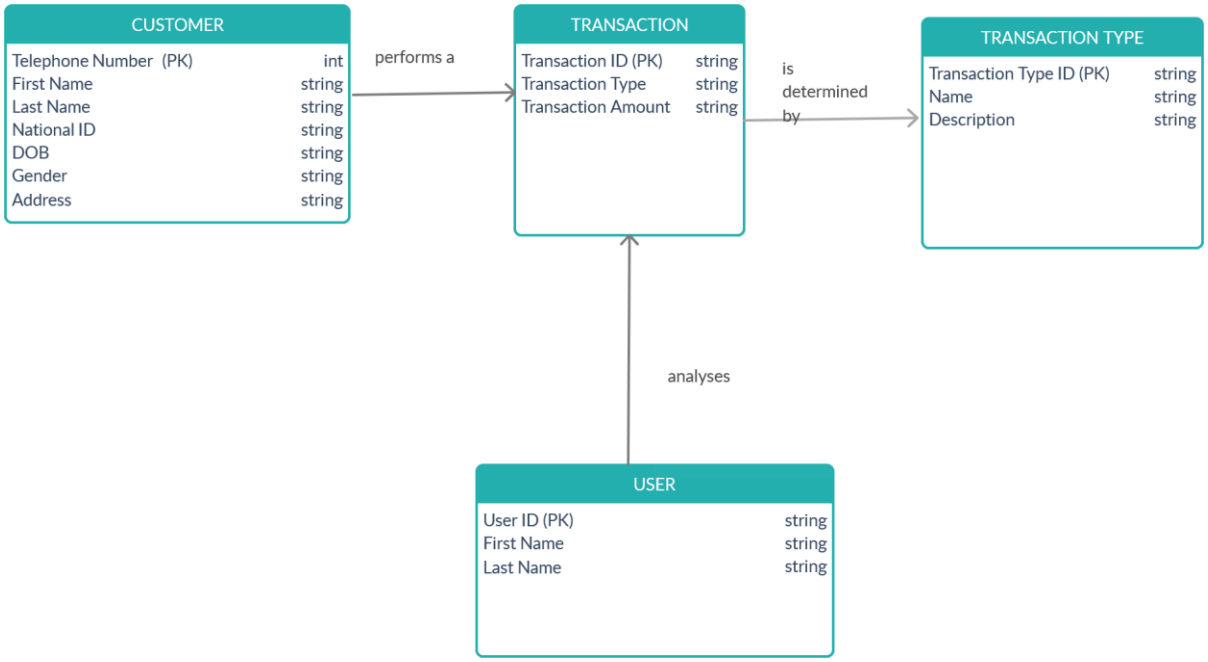


Figure 4.9: Database Schema

Chapter 5: System Development and Testing

5.1 Introduction

This research aimed to develop a tool that would use classification as a pattern recognition technique to actively detect suspicious mobile money transactions. This chapter therefore discusses in detail the process of development and testing of the tool based on functional and non-functional requirements captured in the preceding chapter.

5.2 Detection Model Structure

The proposed Financial crime detection tool applied a classification model to accurately identify suspicious transactions based on unique and differentiating traits. Figure 5.1 describes how the model analyses a transaction up to the point it is classified as suspicious or non-suspicious after which an alert is generated and sent to an authorised user.

Figure 5.1:Financial Crime Detection Model

5.2.1 Importing Transactional Data Source

The model used data from the Paysim Mobile Money Dataset for fraud detection sampled in Appendix 1. The code snippet in figure 5.2 visualizes the dataset and some of its labels which were key in categorizing fraudulent from genuine transactions.

```
[2]: data=pd.read_csv(r"C:\Users\eshiw\PAYSIM DATASET\DATA.CSV")
     data.head()
```

Out[2]:

	step	type	amount	nameOrig	oldbalanceOrig	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0	0
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0	0
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	1	0
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0	0

Figure 5.2: Importing Data

5.2.2 Data Processing

The data processing stage involved compressing the data to capture only what was relevant as for the feature extraction phase. In figure 5.3 the researcher established that the column labels step, type, amount, oldbalanceOrig, newbalanceOrig, oldbalanceDest and newbalanceDest in the dataset were most critical in differentiating fraudulent from genuine transactions.

```
In [5]: x=a.drop(["nameOrig","nameDest","isFraud","isFlaggedFraud"],axis=1)
        x.head()
```

Out[5]:

	step	type	amount	oldbalanceOrig	newbalanceOrig	oldbalanceDest	newbalanceDest
2	1	TRANSFER	181.00	181.0	0.0	0.0	0.00
3	1	CASH_OUT	181.00	181.0	0.0	21182.0	0.00
15	1	CASH_OUT	229133.94	15325.0	0.0	5083.0	51513.44
19	1	TRANSFER	215310.30	705.0	0.0	22425.0	0.00
24	1	TRANSFER	311685.89	10835.0	0.0	6267.0	2719172.89

```
In [8]: scale=StandardScaler()
        x_to_scale=np.array(pd.DataFrame(x,columns=["amount","oldbalanceOrig","newbalanceOrig","oldbalanceDest","newbalanceDest"]))
        X_scaled=scale.fit_transform(x_to_scale)
        X_scaled
```

Out[8]:

```
array([[-0.76826718, -0.25861066, -0.15736583, -0.55147855, -0.64537163],
       [-0.76826718, -0.25861066, -0.15736583, -0.54364917, -0.64537163],
       [-0.11369759, -0.19101914, -0.15736583, -0.54959975, -0.62726484],
       ...,
       [-0.3902935 , -0.18820055, -0.15736583, -0.31814059, -0.37694395],
       [ 0.03079507, -0.17906721, -0.15736583,  0.13139809,  0.10231815],
       [-0.38980716,  1.88206432,  2.11288103, -0.37245863, -0.4285382 ]])
```

Figure 5.3: Data Compressing

5.2.3 Feature Extraction

The feature extraction process involved categorizing transactions into two, ‘isFraud’ (suspicious) and ‘nonfraud’ (genuine) as in the code snippet in figure 5.4.

```
#
frauds = df[df.isFraud == 1]
non_frauds = df[df.isFraud == 0]

frauds['balanceChange'] = frauds['newbalanceOrig'] - frauds['oldbalanceOrig']
non_frauds['balanceChange'] = non_frauds['newbalanceOrig'] - non_frauds['oldbalanceOrig']

frauds_mean_balancechange = frauds['balanceChange'].mean()
nfrauds_mean_balancechange = non_frauds['balanceChange'].mean()

width = 0.8
fig, ax = plt.subplots(1,1, figsize = (10, 6))
ax.bar(0.5, nfrauds_mean_balancechange, width, label='Avg. Non Fraud Account Balance Change', align='center')
ax.bar(1.5, frauds_mean_balancechange, width, label='Avg. Fraud Account Balance Change')
fig.legend(loc='best')
plt.axis([0, 2, -1500000, 200000])
```

Figure 5.4: Feature Extraction

The data was then split into training and test data in the ratio 80:20 respectively.

```
In [36]: X_train,X_test,y_train,y_test=train_test_split(X_final_inner,y,test_size=0.2,random_state=0)

In [37]: print(len(X_train),len(X_test),len(y_train),len(y_test))
         368315 92079 368315 92079

X_train,X_test,y_train,y_test=train_test_split(X_final_inner,y,test_size=0.2,random_state=0)
```

Figure 5.5: Split Data

5.2.4 Training the Model

The K-Nearest Neighbors algorithm was applied to analyse users' transactional behaviour within the transactions and tried to detect suspicious patterns withinin the transactions.

```
In [52]: model=KNeighborsClassifier(n_neighbors=3)

In [53]: model.fit(X_train,y_train)

Out[53]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                              metric_params=None, n_jobs=None, n_neighbors=3, p=2,
                              weights='uniform')
```

Figure 5.6: Training the model

5.3 Testing

Agile testing was adopted while working with agile development methodology for this research. Every iteration had its own testing phase to ensure that any error experienced at each iteration was resolved at that phase.

Testing was conducted at first on the model to validate its detection accuracy and then finally on the tool to validate the its ability to detecting suspicious/fraudulent transactions

5.3.1 Model Testing

5.3.1.1 Confusion Matrix

The confusion matrix was then used to validate the accuracy of the KNN algorithm used in the model.

```
In [20]: print(confusion_matrix(y_test,y_predict))

[[91836    14]
 [   66   163]]
```

Figure 5.7: Confusion Matrix

5.3.1.2 Classification Report

The classification report was run to measure the quality of model in detecting fraudulent transactions.

```
In [21]: print(classification_report(y_test, y_predict))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	91850
1	0.92	0.71	0.80	229
accuracy			1.00	92079
macro avg	0.96	0.86	0.90	92079
weighted avg	1.00	1.00	1.00	92079

Figure 5.8: Classification Report

5.3.1.3 Area Under Receiver Operating Characteristic (ROC)

Accuracy is measure by the area under the ROC curve as in figure 5.8. An area of 1 represents a perfect test, an area of 0.5represents a worthless test. The KNN algorithm achieved an accuracy of 0.85.

```
In [23]: false_positive_rate, true_positive_rate, threshold = roc_curve(y_test, y_predict_proba)
roc_auc = auc(false_positive_rate, true_positive_rate)
plt.title('Receiver Operating Characteristic')
plt.plot(false_positive_rate, true_positive_rate, 'b',
label='AUC = %f'% roc_auc)
plt.legend(loc='lower right')
plt.plot([0,1],[0,1],'r--')
plt.xlim([-0.1,1.2])
plt.ylim([-0.1,1.2])
plt.ylabel('True Positive Rate (TPR=TP/P=TP/(TP+FN))')
plt.xlabel('False Positive Rate (FPR=FP/N=FP/(FP+TN))')
plt.show()
```

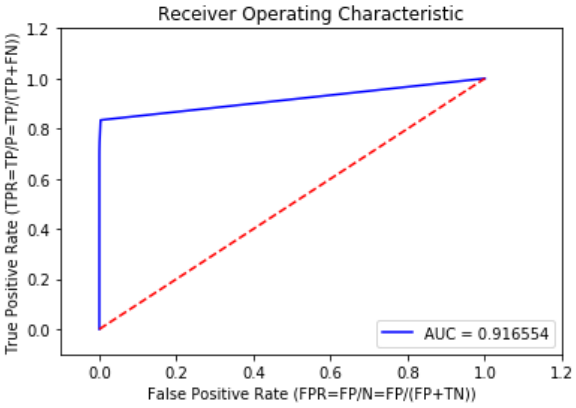


Figure 5.9: ROC Graph

5.3.2 System Testing

The ability of the tool to detect fraudulent transactions was satisfied by applying a set of dummy transactions within a date range of 1st May 2020 to 30th May 2020.

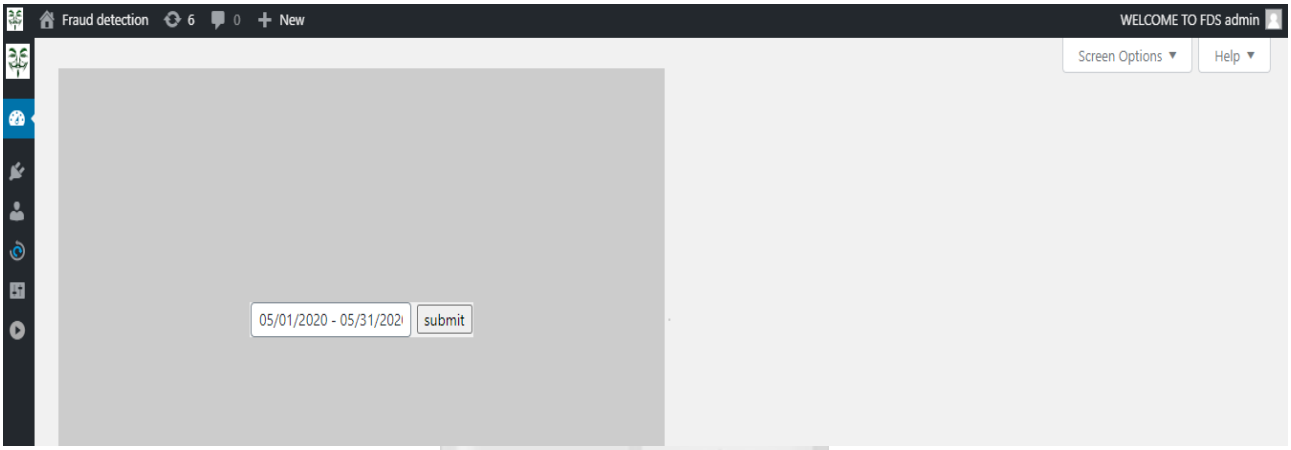
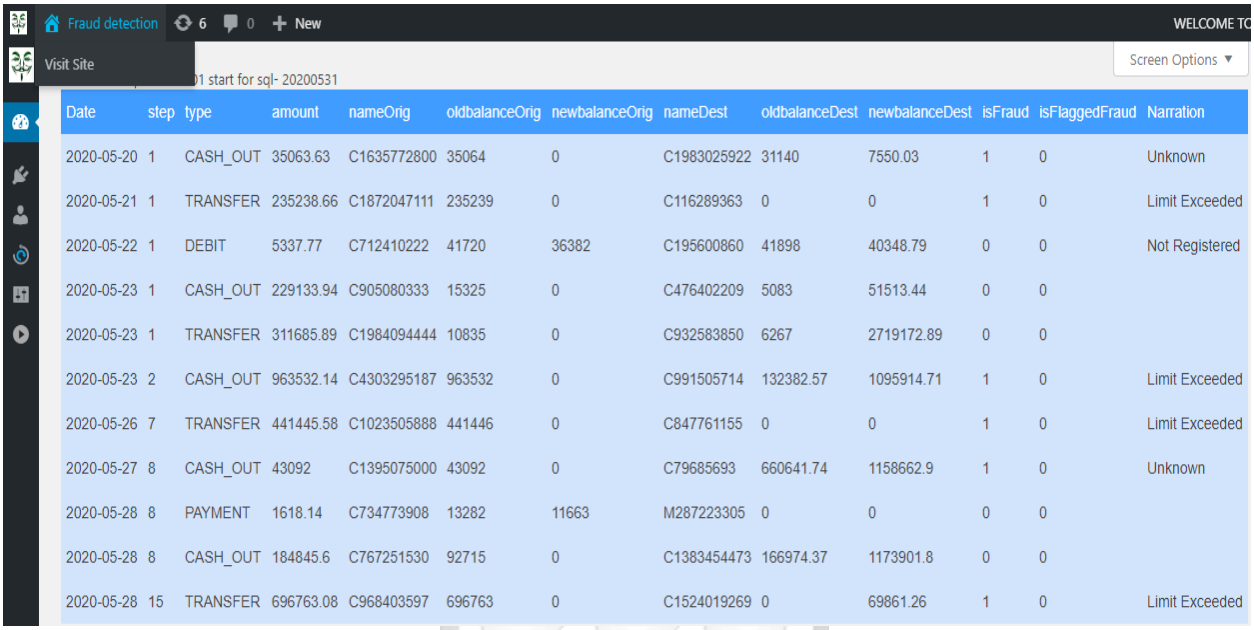


Figure 5.10: Applying date range to filter flagged transactions

The system was able to output the flagged transactions as well as the reasons as to why they were flagged. Transactions were flagged on the basis of transaction limit exceeded or an account not being dully registered.



Date	step	type	amount	nameOrig	oldbalanceOrig	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud	Narration
2020-05-20	1	CASH_OUT	35063.63	C1635772800	35064	0	C1983025922	31140	7550.03	1	0	Unknown
2020-05-21	1	TRANSFER	235238.66	C1872047111	235239	0	C116289363	0	0	1	0	Limit Exceeded
2020-05-22	1	DEBIT	5337.77	C712410222	41720	36382	C195600860	41898	40348.79	0	0	Not Registered
2020-05-23	1	CASH_OUT	229133.94	C905080333	15325	0	C476402209	5083	51513.44	0	0	
2020-05-23	1	TRANSFER	311685.89	C1984094444	10835	0	C932583850	6267	2719172.89	0	0	
2020-05-23	2	CASH_OUT	963532.14	C4303295187	963532	0	C991505714	132382.57	1095914.71	1	0	Limit Exceeded
2020-05-26	7	TRANSFER	441445.58	C1023505888	441446	0	C847761155	0	0	1	0	Limit Exceeded
2020-05-27	8	CASH_OUT	43092	C1395075000	43092	0	C79685693	660641.74	1158662.9	1	0	Unknown
2020-05-28	8	PAYMENT	1618.14	C734773908	13282	11663	M287223305	0	0	0	0	
2020-05-28	8	CASH_OUT	184845.6	C767251530	92715	0	C1383454473	166974.37	1173901.8	0	0	
2020-05-28	15	TRANSFER	696763.08	C968403597	696763	0	C1524019269	0	69861.26	1	0	Limit Exceeded

Figure 5.11: Output of flagged transactions

Chapter 6: Discussion

6.1 Introduction

This main objective of this research was to apply pattern recognition techniques to mobile money transactions for analysis of suspicious patterns and detect financial crimes. This section discusses the study's findings while being guided by the research objectives of the implemented tool.

6.2 Characteristics of mobile money financial transactional activities

The first objective of the research aimed to analyse the characteristics of mobile money transactions that have made them a conduit for transferring their illegally acquired funds undetected. The researcher found that mobile money transactions were grouped as cash deposits, withdrawals, transfers and payments. These transactions collectively are characterized as instantaneous, do not require agent intervention (except for cash deposits or withdrawals) and can be layered multiple transactions in small margins. They are however not anonymous (as implied in numerous literature) in Kenya as all subscribers are required to be registered SIM users before they can transact, in addition every transaction is uniquely identified by a transaction code.

6.3 Effectiveness of financial crimes detection controls and systems

The second objective of the research aimed to analyse the effectiveness of existing systems and controls in financial crime detection. The researcher found that in Kenya ,Financial institutions are mandated by the Central bank of Kenya to have in place fraud detection ,AML and CFT controls based on the FATF AML/CFT guidelines as a risk deterrence and management measure, failure to which attracts hefty fines. Mobile Money Networks are however not listed as traditional financial institutions because the owning company is a Mobile Network Operator. They therefore apply the bare minimum AML/CFT processes which are at customer onboarding and or if an account is flagged for investigation the Criminal investigative unit.

6.4 Research on pattern recognition in detecting financial crimes

The third objective of the research aimed to analyse pattern recognition techniques uses to investigate financial crimes.

The researcher established that pattern recognition was grouped as supervised, unsupervised and semi-supervised. The researcher also found that supervised techniques proved to be most suitable in classifying suspicious transactions due to their simplicity and high accuracy scores in detecting abnormal patterns.

6.5 Design and develop a tool for detecting financial crimes based on pattern recognition in mobile money transactions

The fourth objective of the research aimed to design and develop a pattern recognition tool to detect financial crimes on Mobile money transactions. The researcher designed and develop a supervised pattern recognition tool whose underlying model was the KNN algorithm. This tool was able to detect suspicious transactions from legitimate ones.

6.6 Testing the tool to detect a financial crime on a mobile money platform

The fifth objective of the research was to test the ability of the tool to detect financial crimes on mobile money transactions. The researcher tested the pattern recognition tool to validate its ability that it was able to accurately detect financial crimes using a ROC graph and confusion matrix that yielded an accuracy level of 97%. The tool was therefore determined to be accurate

Chapter 7: Conclusion and Recommendation

7.1 Overview

This research was conducted to analyse the suitability of pattern recognition in the analysis and detection of financial crimes on mobile money transactions. This chapter addresses the study's conclusion based on the research objectives, recommendations, and future work.

7.2 Conclusion

The first objective of the research aimed to analyse the characteristics of mobile money transactions that have made them a conduit for transferring criminals illegally acquired funds undetected. Mobile money transactions were grouped as cash deposits, withdrawals, transfers and payments. These transactions collectively are characterized as instantaneous, do not require agent intervention (except for cash deposits or withdrawals) and can be layered multiple transactions in small margins.

The research further aimed to analyse the effectiveness of existing systems and controls in financial crime detection. The researcher found that in Kenya, unlike traditional Financial institutions, MNOs are not mandated to implement KYC/CFT and anti-fraud controls in totality. Suspected transactions were investigated on an event driven basis and not proactively.

The third objective of the research aimed to analyse pattern recognition techniques used to investigate financial crimes. The researcher established that pattern recognition was grouped as supervised, unsupervised, and semi-supervised. For pattern recognition that involved the processing of labelled data that was voluminous supervised pattern recognition was applied.

The fourth objective of the research aimed to design and develop a pattern recognition tool to detect financial crimes on Mobile money transactions. The researcher designed and developed a supervised pattern recognition tool that was able to detect suspicious transactions from legitimate ones.

The last objective of the research was to test the ability of the tool to detect financial crimes on mobile money transactions. The tool was determined to accurately detect financial crimes on mobile money transactions.

7.3 Recommendations

The potential application of such a tool is limitless especially in areas of improved revenue collection where a tax agency can verify the tax compliance of taxpayers by monitoring how they transact and how much they transact and bringing tax evaders to book. Also, to curb corruption especially in third world countries where mobile money uptake is high and hence used as a channel to fund corruption. Governments can monitor their employees accounts for suspicious transactions.

As new technologies emerge criminals find new ways to attack or carry out suspicious activities while going undetected. It is important that the methods of securing transactions are equally robust if not faster to minimize risk while maintaining regulatory compliance.

Finally, a system is only as efficient as the effective implementation of the Standard Operating Procedures put in place by an organization and CDD play an important role when onboarding customers. Also, continuous customer due diligence is key in ensuring the risk exposure of a firm is kept at a minimum.

7.4 Future Works

In the research proposed to detect financial crimes which is an umbrella that covers fraud, money laundering and terrorist funding. Future research can expound on this area by detecting the exact type of crime based on the transaction threshold set by an MNO per time duration while incorporating the use of sanction lists to establish the risk rating of mobile money users.

GPS analysis can also assist security agencies in apprehending fraudsters before they get away with the crime and algorithms for faster execution should be considered especially considering the high volumes of transactions that are processed at a at time.

References

Adedoyin, A. (2018, June). PREDICTING FRAUD IN MOBILE MONEY TRANSFER. London, United Kingdom.

Aguilar, J. (2004, October). Statistical Pattern Recognition Problems and the Multiple Classes Random Neural Network Model.

Alexandre, C. (2010, November 22). *Business and Markets*. Retrieved from CGAP: https://www.cgap.org/blog/10-things-you-thought-you-knew-about-m-pesa

Anosh, G., & Ahmadi, M. (2015, October 11). Money Laundering and Financing of Terrorism. Mumbai, Mumbai, India.

Asht, S., & Dass, R. (2012, August 8). "“Pattern Recognition Techniques: A Review”. *International Journal of Computer Science and Telecommunications, 2*(8).

Bahia, K., & Muthiora, B. (2019, February). The Mobile Money Regulatory Index. London, United Kingdom.

BENEDETTI, M. (2018, January 24). *The Advantages and Limitations of Synthetic Data*. Retrieved from Samasource: https://www.samasource.com/blog/2018/01/24/the-advantages-and-limitations-of-synthetic-data

Bettcher, k., & Mihaylova, T. (2015, May 26). Economic Inclusion: Leveraging Markets and Entrepreneurship to Extend Opportunity. Washington, Washington, USA: Center for International Private Enterprise .

Bishop, C. (2006). *Pattern Recognition and Machine Learning.* singapore: Springer Science+Business Media.

Brink , P., & Marilynn , W. (1998). Exploratory-Descriptive Designs. CA, Thousand Oaks, USA.

Brownlee, J. (2016, April 1). *Logistic Regression for Machine Learning*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/logistic-regression-for-machine-learning/

Brungs et al. (2008). *Fraud Profiling Conceptual Model.*

Buku, M. W., & Mazer, R. (2017). *Fraud in Mobile Financial Services: Protecting Consumers, Providers,and the System.* Washington DC: CGAP.

Buku, M., & Mazer, R. (2017, april). *Fraud in Mobile Financial Services.* Retrieved from CGAP: http://www.cgap.org/sites/default/files/Brief-Fraud-in-Mobile-Financial-Services-April-2017.pdf

Caputo, S. (2019, August 20). *Mobile for Development:The pivotal role of mobile money agents in driving financial inclusion*. Retrieved from GSMA: https://www.gsma.com/mobilefordevelopment/blog/the-pivotal-role-of-mobile-money-agents-in-driving-financial-inclusion/

CBK. (2011). THE NATIONAL PAYMENT SYSTEM ACT. Nairobi, Kenya.

Chartis Research. (2015, April). Financial Crime Risk Management Systems Oracle Vendor Highlights. London, United Kingdom.

Chatain, P., Hernández-Coss, R., & Borowik, K. (2008). Integrity in Mobile Phone Financial.

Chatain, P.-L., Zerzan, A., Noor, W., Dannaoui, N., & Koker, L. (2011). *Protecting Mobile Money against Financial Crimes.* Washington: The World Bank. doi:10.1596/978-0-8213-8669-9

Choudhary, P. (2017, February 15). *Data science*. Retrieved from Oracle Data science blog: https://blogs.oracle.com/datascience/introduction-to-anomaly-detection

Luo, X. (2014). *Suspicious Transaction Detection for Anti-Money Laundering.*

Colbert, E. (1990). The object-oriented software development method: a practical approach to object-oriented development. *Tri-Ada '89*, (pp. 400–415). CA.

Comply Advantage. (2018, July 31). *Transaction Monitoring*. Retrieved from https://complyadvantage.com: https://complyadvantage.com/knowledgebase/anti-money-laundering/transaction-monitoring/

Croall, H. (2005). White Collar Crime. *Transnational and Comparative Criminology Glasshouse Press*, 227-246.

Davidson, N. (2012, March). Mapping and Effectively Structuring Operator-Bank Relationships to Offer Mobile Money for the Unbanked. *Mobile Money for the Unbanked*. London, London, United Kingdom: GSMA.

De, U. (2019, November 10). *Types of Distances in Machine Learning*. Retrieved from Analytics Vidhya: https://medium.com/analytics-vidhya/types-of-distances-in-machine-learning-5b1233380775

Didimo, W., & Liotta, G. (2014, August). Network visualization for financial crime detection. *Journal of Visual Languages & Computing, 25*(4).

Duin, R., Roli, F., & Ridde, D. (2002). A note on core research issues for statistical. *Pattern Recognition Letters*, 493–499.

Ernst & Young. (2011). Mobile Money. United Kingdom.

Ernst and Young. (2016). Effective screening for sanctions and AML Risk Management. London, United Kingdoom. Retrieved from Academia.EDU.

Field, R. (2012, August 22). *Safaricom strengthens MPESA security*. Retrieved from Critical Analysis for Telecommunications Executives: https://www.commsmea.com/12561-safaricom-strengthens-mpesa-security

Financial Action TAsk Force FATF*GAFI. (2010). *An introduction to the FATF and its work*. Retrieved from Financial Action TAsk Force FATF*GAFI: https://www.fatf-gafi.org/media/fatf/documents/brochuresannualreports/Introduction%20to%20the%20FATF.pdf

Financier WorldWide. (2018). *Managing financial crime risk and AML processes with technology.* Lichfield: Financier WorldWide.

FinScan. (2016, June 12). *AML Compliance:*. Retrieved from Innovative Systems: https://www.innovativesystems.com/finscan-aml-compliance

Fisher, D. (2017, December 11). Customer Due Diligence Process. London, London, United Kingdom.

Franzese, M., & Luliano, A. (2018). *Encyclopedia of Bioinformatics and Computational Biology.* Sydney.

Flovic, V. (2018, December 31). *Machine Learning*. Retrieved from Towards DataScience: https://towardsdatascience.com/how-to-use-machine-learning-for-anomaly-detection-and-condition-monitoring-6742f82900d7

Gao, Z., & Ye, M. (2007). A framework for data mining-based anti-money laundering research. *Journal of Money Laundering Control, 10*(2), 170-179. doi:https://doi.org/10.1108/13685200710746875

Gonzalez, R. (2008). *Object Recognition in Digital image processing* (3rd ed.). Pearson.

GSMA. (2017). *State of the Industry Report on Mobile Money.* London: GSMA.

GSMA. (2018). State of the Industry Report on Mobile Money. London, London, United Kingdom.

Gujral , S., Alam , M., Bhavsar , J., & Khan , S. (2019, November). Pattern Recognition.

Gunal, D. (2008, October). "AUTOMATED CATEGORIZATION SCHEME FOR DIGITAL LIBRARIES IN DISTANCE LEARNING:A Pattern Recognition Approach,". *Turkish Online Journal of Distance Education-TOJDE, 9*, 4.

Guyon, I., & Elisseeeff, A. (2003). *An Introduction to Variable and Feature Selection.* Andr´e Elisseeff.

Harrison, O. (2018, September 10). *Machine Learning*. Retrieved from towards data science: https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761

MONITOR FOR DETECTING MONEY LAUNDERING AND TERRORIST FINANCING. *Journal of Theoretical and Applied Information Technology, 85, 3*. Cairo, Egypt: JATIT & LLS.

Hossam, T., Mohamed, Z., Tarek, S., & Khaled, B. (2016). Design of a Monitor for Detecting Money Laundering and Terrorist . *Journal of Theoretical and Applied Information* , 425-436.

Jain, Duin, & Mao. (2000). Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell*, 4-37.

Jun, T. (2006). A Peer Dataset Comparison Outlier Detection Model Applied to Financial Surveillance. *Proceedings of the 18th International Conference on Pattern Recognition*. Washington, USA: IEEE Computer Society.

Jung, J., & Lee, J. (2017). Contemporary Financial Crime. *Journal of Public Administration and Governance, 7*(2), 88-97. doi:10.5296/jpag.v7i2.11219

Kannan, S., & Somasundram, K. (2017). Autoregressive-based outlier algorithm to detect money laundering activities. *Journal of Money Laundering Control*.

Kiplagat, S. (2020, February 28). *Companies*. Retrieved from Business Daily: https://www.businessdailyafrica.com/corporate/companies/Safaricom--data-firm-in-M-Pesa-staff-fraud-fight/4003102-5471712-154i6jjz/index.html

KPMG. (2019). *Combating financial crime:*. Retrieved from KPMG:Insights: https://home.kpmg/xx/en/home/insights/2019/03/combating-financial-crime-fs.html

Lal, R., & Sachdev, I. (2015, July). Mobile Money Services - Design and Development for Financial Inclusion. Massachusetts, Boston, United States.

Liu, J., SUn, J., & Wang, S. (2006, June). "Pattern Recognition: An overview,". *IJCSNS International Journal of Computer Science and Network Security, 6*.

Lopez-Rohas, E. (2016). PAYSIM: A FINANCIAL MOBILE MONEY SIMULATOR FOR FRAUD DETECTION. *The 28th European Modeling and Simulation Symposium-EMSS.* Larnaca.

Lopez-Rojas, E. (2016). *APPLYING SIMULATION TO THE PROBLEM OF DETECTING FINANCIAL FRAUD.* Blekinge Institute of Technology, Department of Computer Science and Engineering. Karlskrona: Blekinge Institute of Technology.

Luell, J. (2010). *Employee Fraud Detection under Real World Conditions.* Zurich: University of Zurich.

Mahindra Comviva. (2016, May 16). *The basics of Mobile Money Security*. Retrieved from Comviva: https://blog.comviva.com/the-basics-of-mobile-money-security/

MCA, & Prakaran. (2014). *Multi-Variant Relational Model for Money Laundering Identification using Time Series Data Set architecture.*

McGrath, F., & Lonie, S. (2013). Platforms for Successful. London, United Kingdom.

Mehmet et al. (2013). *Money laundering detection framework to link the disparate and evolving schemes.*

Merton, C. (1990). The Financial System and Economic Performance . *Journal of Financial Services Research*, 263-300.

Mule, J. (2015, November 23). *What does M-PESA use for security for its financial transaction system?* Retrieved from Quora.com: https://www.quora.com/What-does-M-PESA-use-for-security-for-its-financial-transaction-system

Mutai, E. (2019, March 29). *News*. Retrieved from Daily Nation: https://www.nation.co.ke/news/MPs-seek-CBK-regulation/1056-5047788-2wdi1l/index.html

Ombaka, R. (2018, October 2). Safaricom, Airtel, Telkom ease mobile money transfer across networks. Nairobi, Nairobi, Kenya.

Omondi, D. (2019, January 29). *Financial Standard*. Retrieved from Standard Digital: https://www.standardmedia.co.ke/article/2001311163/mobile-cash-transfers-pose-new-headache-for-security-agencies-local-banks

Oracle. (2018). Addressing Financial Crime and Compliance Challenges with Advanced Analytics.

Owuor, J. (2016). *Anti- Money Laundering & Combating Terrorism Financing.* Nairobi: Insurance Regulatory Authority Kenya.

Perachio, G. (2017, October). The Future of Trader Surveillance. London, United Kingdom.

Perelman, N. (2019, October). *Machine Learning*. Retrieved from Anodot: https://www.anodot.com/blog/deliver-results-scale-supervised-vs-unsupervised-anomaly-detection/

Privacy International. (n.d.). *Topics: SIM Card Registration*. Retrieved from Privacy International: https://privacyinternational.org/topics/sim-card-registration

Raj, M., Saini, J., & Parmar, D. (2015). Applications of Pattern Recognition Algorithms in Agriculture: A Review. *International Journal of Advanced Networking and Applications*, 2495-2502.

Rudin, C., & Letham, B. (2013). Growing a list. Data Mining and Knowledge Discovery .

Sayantika, B. (n.d.). *Money:Money in Economics*. Retrieved from Micro Economic Notes : http://www.microeconomicsnotes.com/money/money-in-economics-definition-types-functions-characteristics-importance-and-evils-economics/15100

Senator, T., Goldberg, H., Wooton, J., Cottini, M., & Wong, R. (1995). Identifying Potential Money Laundering from Reports of large cash transactions. *The FinCEN Artificial Intelligence System, 16(4)*, 156-170. USA: U.S. Department of the Treasury - Financial Crimes Enforcement Network (FinCEN), Association for the Advancement of Artificial Intelligence.

Serratosa, F., ALQUÉZAR , R., & SANFELIU, A. (2011, November 21). SYNTHESIS OF FUNCTION-DESCRIBED GRAPHS AND CLUSTERING OF ATTRIBUTED GRAPHS. *International Journal of Pattern Recognition and Artificial Intelligence, 16*(6). Retrieved from World Scientific: https://www.worldscientific.com/doi/abs/10.1142/9789814343138_0008

Sharma, & Panigrahi. (2013). *Data Mining Framework for Financial Accounting Fraud Detection* .

Sharma, P., & Kaur, M. (2013, April). Classification in Pattern Recognition. *International Journal of Advanced Research in Computer Science and Software Engineering: A Review, 3*(4).

Solin, & Zerzan. (2010). *GSMA Risk Assesment Methodology.* GSMA.

Stackify. (2017, september 17). *Agile-Methodolog*. Retrieved from Stackify: https://stackify.com/agile-methodology/

Subia, M., & Martinez, N. (2014). Overview and Opportunities. *mobile money services:"A bank in your pocket".* ACP Observatory on Migration.

Suri, T., & Jack, W. (2011). Mobile Money Dataverse. Columbia, Washington, USA.

Tang, J., Alelyani, S., & Liu, H. (2014, January 1). Feature Selection for Classification. *Data Classification: Algorithms and Applications*, pp. 37-64.

Theodorou, Y., Okong'o , K., & Yongo, E. (2019). *Access to Mobile Services and Proof of Identity 2019:Assessing the impact on digital and financial inclusion.* London: GSMA.

Tookitaki. (2019, January 14). *Transaction Monitoring*. Retrieved from Tookitaki: https://www.tookitaki.ai/news-views/mas-stress-on-transaction-monitoring-for-effective-aml-cft-compliance-and-machine-learning-is-the-answer/

UNCTAD. (2018, December 13). Mobile money holds key to financial inclusion in Africa. Geneva, Switzerland: United Nations Conference on Trade and Development.

Valchev, M. (2019, July 29). *Mobile Wallet*. Retrieved from Software Group: https://www.softwaregroup.com/insights/blog/individual-article/main-blog/2019/08/01/15-key-features-that-make-your-mobile-wallet-stand-out

Wang, S., & Yang, J. (2007). A money laundering risk evaluation method based on decision tree. *Proceedings of the International Conference on Machine Learning and Cybernetics*, 283-286. Hong Kong.

Wang, T., & Rudin, C. (2013, August). Learning to Detect Patterns of Crime. Cambridge, Massachussets, USA.

Wang, Y., Wang, H., Gao, S., & Xu, D. (2008). Intelligent Money Laundering Monitoring and Detecting System. *European and Mediterranean Conference on Information Systems*. Dubai.

Weber, M. (2018, November 30). Scalable Graph Learning for Anti-Money Laundering: A First Look.

Wells , C. J. (2009, 01 28). *Development Methodologies*. Retrieved 2020, from www.technologyuk.net: http://www.technologyuk.net/computing/software-development/systems-analysis/methodologies.shtml

WGBIS. (2014, december 29). *"Lesson 1: Image Processing and Interpretation*. Retrieved from Environmental Information System (ENVIS): http://wgbis.ces.iisc.ernet.in/envisrs/?q=node/26/

Wolfsberg Group. (2019, March). Wolfsberg Guidance on AML and CFT. north-eastern, Château Wolfsberg, Switzerland.

Zetzsche, Dirk, & Arner. (2018). Digital ID and AML/CDD/KYC utilities for financial inclusion, integrity and competition.

Appendix 1: Portion of Paysim Synthetic Dataset for Fraud Detection

step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
1	PAYMENT	9839.64	C1231006815	170136	160296.36	M1979787155	0	0	0	0
1	PAYMENT	1864.28	C1666544295	21249	19384.72	M2044282225	0	0	0	0
1	TRANSFER	181	C1305486145	181	0	C553264065	0	0	1	0
1	CASH_OUT	181	C840083671	181	0	C38997010	21182	0	1	0
1	PAYMENT	11668.14	C2048537720	41554	29885.86	M1230701703	0	0	0	0
1	PAYMENT	7817.71	C90045638	53860	46042.29	M573487274	0	0	0	0
1	PAYMENT	7107.77	C154988899	183195	176087.23	M408069119	0	0	0	0
1	PAYMENT	7861.64	C1912850431	176087.23	168225.59	M633326333	0	0	0	0
1	PAYMENT	4024.36	C1265012928	2671	0	M1176932104	0	0	0	0
1	DEBIT	5337.77	C712410124	41720	36382.23	C195600860	41898	40348.79	0	0
1	DEBIT	9644.94	C1900366749	4465	0	C997608398	10845	157982.12	0	0
1	PAYMENT	3099.97	C249177573	20771	17671.03	M2096539129	0	0	0	0
1	PAYMENT	2560.74	C1648232591	5070	2509.26	M972865270	0	0	0	0
1	PAYMENT	11633.76	C1716932897	10127	0	M801569151	0	0	0	0
1	PAYMENT	4098.78	C1026483832	503264	499165.22	M1635378213	0	0	0	0
1	CASH_OUT	229133.9	C905080434	15325	0	C476402209	5083	51513.44	0	0
1	PAYMENT	1563.82	C761750706	450	0	M1731217984	0	0	0	0
1	PAYMENT	1157.86	C1237762639	21156	19998.14	M1877062907	0	0	0	0
1	PAYMENT	671.64	C2033524545	15123	14451.36	M473053293	0	0	0	0
1	TRANSFER	215310.3	C1670993182	705	0	C1100439041	22425	0	0	0
1	PAYMENT	1373.43	C20804602	13854	12480.57	M1344519051	0	0	0	0

Appendix 2: Ethical Approval

Strathmore UNIVERSITY

25th February 2020

Ms Eshiwani, Michelle
michelle.eshiwani@strathmore.edu

Dear Ms Eshiwani,

RE: Detecting Financial Crimes using Pattern Recognition Techniques

This is to inform you that SU-IERC has reviewed and **approved** your above research proposal. Your application approval number is **SU-IERC0636/20**. The approval period is **25th February, 2020 to 24th February, 2021.**

This approval is subject to compliance with the following requirements:

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

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

Yours sincerely,

for: Dr Virginia Gichuru,
Secretary; SU-IERC

STRATHMORE UNIVERSITY INSTITUTIONAL ETHICS REVIEW COMMITTEE
SU-IERC
25 FEB 2020
TEL: +254 (0)703 034 000
P. O. Box 59857 - 00200
NAIROBI - KENYA

Cc: Prof Fred Were,
Chairperson; SU-IERC

Appendix 3: Research License

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION

Ref No: 787909

Date of Issue: 18/March/2020

RESEARCH LICENSE

This is to Certify that Ms.. Michelle Mercy Eshiwani of Strathmore University, has been licensed to conduct research in Nairobi on the topic: DETECTING FINANCIAL CRIMES USING PATTERN RECOGNITION TECHNIQUES for the period ending : 18/March/2021.

License No: NACOSTI/P/20/4135

787909

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

Appendix 4: Turnitin Report

Turnitin Originality Report

Processed on: 09-Apr-2020 9:32 AM EAT
ID: 1286903699
Word Count: 13561
Submitted: 9

Detecting Financial Crimes using Pattern Recognition Techniques By Michelle Mercy Eshiwani

Similarity Index: 26%

Similarity by Source
Internet Sources: 17%
Publications: 9%
Student Papers: 18%

1% match (student papers from 03-May-2019) Submitted to University of the Western Cape on 2019-05-03
1% match (Internet from 13-Mar-2017) http://www.virtusinterpress.org/IMG/pdf/JGB__Volume_4_Issue_4_2015_Continued6_-3.pdf
1% match (Internet from 08-Apr-2020) https://www.standardmedia.co.ke/article/2001311163/mobile-cash-transfers-pose-new-headache-for-security-agencies-local-banks
1% match (Internet from 17-Jun-2017) http://www.i-scholar.in/index.php/TIANA/article/download/140864/129133
1% match (Internet from 03-Nov-2014) http://www.globalinitiative.net/download/financial-crime/global/World%20Bank%20-%20Protecting%20Mobile%20Money%20against%20Financial%20Crimes.pdf
1% match (Internet from 07-Apr-2020) https://www.samasource.com/blog/2018/01/24/the-advantages-and-limitations-of-synthetic-data
1% match (publications) Cross Gombiro, Mmaki Jantjies, Nehemiah Mavetera, "A CONCEPTUAL FRAMEWORK FOR DETECTING FINANCIAL CRIME IN MOBILE MONEY TRANSACTIONS", Journal of Governance and Regulation, 2015
1% match (student papers from 27-Jul-2017) Submitted to Jawaharlal Nehru University (JNU) on 2017-07-27
1% match (Internet from 16-Aug-2019) https://www.financierworldwide.com/forum-managing-financial-crime-risk-and-aml-processes-with-technology
1% match (Internet from 08-Jul-2019) http://csse.xjtlu.edu.cn/csseWiki/GangminLiProjectPage
< 1% match (Internet from 24-Mar-2016) http://www.ijcna.org/Manuscripts%5CVolume-1%5CIssue-1%5CVol-1-Issue-1-M-03.pdf
< 1% match (publications) Rehema Kagendo Kiarie, "chapter 9 Ethics in Mobile Banking", IGI Global, 2020
< 1% match (Internet from 22-Dec-2017) https://issuu.com/world.bank.publications/docs/9780821386699
< 1% match (publications) Zulfiqar Ali, Syed Khurram Shahzad, Waseem Shahzad, "Performance Analysis of Statistical Pattern Recognition Methods in KEEL", Procedia Computer Science, 2017
< 1% match (Internet from 27-Nov-2019) https://www.worldscientific.com/doi/pdf/10.1142/9789814343138_0008
< 1% match (student papers from 17-Aug-2015) Submitted to Vrije Universiteit Brussel on 2015-08-17
< 1% match (Internet from 26-Dec-2019) https://www.emerald.com/insight/content/doi/10.1108/JMLC-07-2016-0031/full/html
< 1% match (Internet from 20-Dec-2019) https://pt.scribd.com/document/52943757/Protecting-Mobile-Money-against-Financial-Crimes
< 1% match (Internet from 25-Nov-2019) https://rd.springer.com/article/10.1007/s10044-013-0322-1
< 1% match (Internet from 28-Sep-2019) http://www.technologyuk.net/computing/software-development/systems-analysis/methodologies.shtml
< 1% match (student papers from 08-Oct-2012) Submitted to North West University on 2012-10-08
< 1% match (Internet from 02-Sep-2019) http://talisman-intl.com/background-of-inventory-system.html
< 1% match (student papers from 16-Mar-2015) Submitted to Yeditepe University on 2015-03-16
< 1% match () http://hdl.handle.net/10948/d1020079