DEVELOPMENT OF A WEB-BASED MACHINE LEARNING MONEY LAUNDERING DETECTION AND PREVENTION MODEL

BY

HAMPO, JOHNPAUL ANENECHUKWU CHUKWUNONSO

JULY, 2024.

DEVELOPMENT OF A WEB-BASED MACHINE LEARNING MONEY LAUNDERING DETECTION AND PREVENTION MODEL

BY

HAMPO, JOHNPAUL ANENECHUKWU CHUKWUNONSO
(B.Sc.)
20164023558

A THESIS SUBMITTED TO
THE POSTGRADUATE SCHOOL
FEDERAL UNIVERSITY OF TECHNOLOGY, OWERRI

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE (M.SC) DEGREE IN COMPUTER SCIENCE.

JULY, 2024.

CERTIFICATION

This is to certify that this work "Development of a Web-based Machine Learning Money Laundering Detection and Prevention Model", was carried out by **Hampo, JohnPaul Anenechukwu Chukwunonso (20164023558)** in partial fulfilment for the award of Master of Science (M.Sc.) degree in Computer Science in the Department of Computer science, School of Information and Communication Technology, Federal University of Technology, Owerri.

Dr. (Mrs) E. C. Nwokorie
Supervisor

22/01/2025
Date

Dr. (Mrs) J. N. Odii
Co-Supervisor

23/01/2025
Date

Dr. (Mrs) J. N. Odii
Head of Department

23/01/2025
Date

Prof. (Mrs) U. F. Eze
Dean, SICT

23/01/25
Date

Prof. (Mrs.) J. N. Nwosu
Dean, Postgraduate School

Date

Prof. Moses Okechukwu Onyesolu
External Supervisor

Date

DEDICATION

This research is dedicated to the memory of my father, Mr. Hampo, A. Cyprian.

ACKNOWLEDGEMENTS

I am grateful to my supervisors Dr (Mrs.) E. C. Nwokorie and Dr (Mrs.) J. N. Odii for their guidance, suggestions and corrections to this work.

I will not fail to acknowledge all my lecturers: Prof. E. N. Erumaka, Prof. Aloy C. Onyeka, Prof. E.O. Nwachukwu, Prof. Prince Asagba, Dr. (Mrs.) Juliet N. Odii, Dr. Celestine Njoku, Mr. Joseph I. Eke, Dr. Stanley A. Okolie, Dr. Jacinta C. Odirichukwu, Dr. (Mrs.) Uchenna C. Onyemauche, Dr. (Mrs.) Mercy E. Benson-Emenike, Mr. Stanley O. Diala, Dr. Chinwe G. Onukwugha, Dr. Chukwuma D. Anyiam, Dr. (Mrs.) Chidimma I. Okpalla, Mr. Godson E. Ahamba, Dr. Francisca O. Nwokoma, Dr. Donatus O. Njoku, Dr. Kelechi A. Douglas, Mr. Peter K. Joseph, Mr. Friday J. Oforma, Mrs. Ezi Achama Eke, Mr. Chukwudi N. Akujobi, Mrs. Idu E. Iheukwumere, Mr. Ikechukwu Onyeanu, Mrs. Stella Egwu-Ahaotu, Dr. C. P. Oleji in Computer Science Department and the able Dean of SICT, Prof. (Mrs.) U. F. Eze for their educational and moral support throughout my programme.

Finally, I sincerely appreciate my mother - Mrs Maria O. Hampo, siblings and in-laws who encouraged and supported me at diverse times throughout this programme and thesis.

I express my heartfelt thanks to all my colleagues notably Mr. Charles, Mr Kanu and Mrs Faith.

Above all, I thank Yahweh for His mercies and grace upon my life and educational journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Explicitly programmed systems, rule-based systems and machine learning systems exist as anti-money laundering systems, however, these systems are for the detection and not prevention of money laundering. This thesis is concerned with the detection and prevention of money laundering by developing a web-based model that uses machine learning (ML) to detect and prevent money laundering transactions. Money laundering which is synonymous with clothes laundering is the process of transforming the real nature of the source of income or money which is usually an illegitimate source to a legitimate source. The model was developed using open datasets on financial transactions from Kaggle.com, which is an open-source website that holds a lot of data. Questionnaires were administered for data acquisition and requirement collection. The questionnaire was given out to people in the banking sector, and the data were analysed to reveal that most respondents see a need for this system and believe it will lead to better financial monitoring and decision-making. The RAD software methodology was applied and Python programming language and Python frameworks were used for this model. recall of 100%, an f1 score (f-measure) of 99.2% and a precision of 98.3% were achieved by this research against the existing system's metrics of 97%, 97% and 98% for f1-score, precision and recall respectively. Also, an accuracy of 98.4% and 81.9% was achieved for the detection model and the prevention model respectively.

Keywords: Machine Learning Techniques; kNN; Prediction; Money Laundering; Classification, kNeighbours classifier

CHAPTER ONE

INTRODUCTION

1.1 Background Information

The world today has drastically changed from the world of last decades. These changes are in areas that are dependent on technology and it is due to the advancement and growth in technology as a result of the birth of state-of-the-art technological ideas, algorithms and systems. These changes are both good but with a resultant bad effect due to the presence of negative minds. The changes are technologically driven (Lokanan, 2022).

Money which is a means of exchange either for goods purchased or for services rendered, is very important in the society and to human. Money is meant to be got legally, but criminals and the notorious minds illegally get money (Doppalapudi et al., 2022). When money is not properly (legally and legitimately) got, it becomes difficult for it to be spent due to the presence of enacted laws, law enforcement agencies/agents and the financial regulatory bodies. Some sources from which dirty money is gotten from are destruction of Critical National Infrastructure (CNI), trafficking, drugs and kidnapping.

The possessors of illegal money need a way to spend their money without being detected by the law enforcement agents and financial regulatory bodies. Thus, criminals and notorious minds tend to make illegal money legal. This is done through some processes and it is termed money laundering. A closer look at the businesses that fully encompass the economy these days, appears to be funded through illegitimate sources despite from afar, they appear to be funded through legitimate sources (Ramya et al., 2022).

Money laundering is the process of transforming the real nature of the source of an income or money which is usually an illegitimate source to a legitimate source. In simple terms, it is making illegal money (money from bad sources) to become legal money (money from good sources). Money laundering is defined as "*an activity that knowingly engages in a financial transaction with the proceeds of some unlawful activity with the intent of promoting or carrying-on that unlawful activity to conceal or disguise the nature, location, source, ownership or control of these proceeds*" (Rafay, 2021).

Money laundering which entails processes of making dishonest and prohibited proceeds appear honest and accepted is an international criminal doing, and it is present in different geographical entities. Money laundering is inter-location and intra-location and the advancement of the Internet and communication technologies has not helped to reduce the vast nature of money laundering in terms of geographical landscape.

An illustration for better understanding is thinking of money as clothes. The clothes can be clean or dirty. The clean clothes can be used by a sane and responsible person without anyone raising an eye or questioning, so is legal ('clean') money. Conversely, when a sane and responsible person put on a dirty cloth, people will raise an eye and there will be lots of questioning, and so is it with illegal ('dirty') money. For the dirty clothes to be used, they have to be laundered by soaking, washing, drying and ironing. These processes are also in money laundering but with different terminologies, as they turn illegal money to a legal money (Dalpiaz, 2020).

The above illustration suggests that money laundering is indulged in by responsible and highly respected people but they use other people, usually not connected to them in blood; to carry out their activities. This is to cover their identity thereby making the process of discovering them hard.

The processes involved in money laundering are: Collection, Placement, Layering and Integration. Criminals need a legal investment to cover-up the criminal operations and the proceeds from criminal operations if not covered up, will be imperilled to investigation, apprehension and forfeiture if need be; hence money laundering to savage both the criminals and their proceeds (OECD, 2019)

Money laundering is not a new crime or phenomenon in the financial world (OECD, 2013). Though the exalt origin cannot be determined. However, literatures stated that its origin is traced to the Chinese traders and the mafia owners of Laundromats, United States (Kingston, 2020; Rafay, 2021). Laundromat's owners needed to prove their legitimacy of their monies due to their earnings from extortion, gambling, bootleg liquor and prostitution which are some of the illicit businesses.

The global estimated value of money laundering in a year was given by the United Nations office on drugs and crime to range from $500 billion to $1 trillion (Cem, 2023; UNODC, 2011). This high range is due to the camouflage done by the practitioners of money laundering. This evil and devious act termed money laundering is done internationally and locally mostly by those the society see as responsible. While criminal is the word used for committers of crime, money launder is used for those involved in money laundering.

Every crime affects the society negatively and so is it with money laundering? The effects are seen not only in the economy but also in the security and political domain of the nation. Also, it discourages diligent and hardworking culture among the populace. Authors like (Enofe et al., 2018) and (Rafay, 2021) observed that money laundering has a devastating effect on the financial health of the citizens of developing nations and that money laundering is on the rise as observed

by the United Nations due to reasons like greed, societal problems, ineffectiveness of prosecuting laws, likelihood of absconding from the stated punishment by law.

Anti-money laundering laws are enacted in Nigeria and other countries of the world, so as to crisscross the devilish crime called money laundering. Nevertheless, the ineffectiveness of executing these laws both at national and international level has put money laundering as the third major international business (actually a business crime) (Le-Nhien et al., 2010; Soltani et al., 2016; UNODC, 2011) with an estimated value range of $500 billion to $1 trillion per year as reported by the United Nations office on drugs and crime in their report (UNODC, 2017). The causation factors to the ineffectiveness of money laundering enacted laws are the application of traditional anti-money laundering approach (technique) and manual investigation systems, which devours time and resources of the anti-money laundering custodian agent.

The effects of money laundering are on the country's economy, security, image and politics, as well as the individuals and organisations that are not into money laundering. Nation's financial stability and international security are preys of money laundering (Soltani et al., 2016). To this effect, laundering of money should not be welcomed by nations. However, the approach used by nations that have not incorporated machine learning in the issues of money laundering is the encouragement of money laundering in disguise.

Data is on the increase daily and there is need for space; this is due to increase in the connectivity and usage of different social media platforms. Systems that are programmed explicitly cannot handle the category of data in this time due to data's velocity, veracity, volume and variety. Hence the need for data mining techniques such as machine learning which implicitly programs system to learn from experience and improves its performance.

Machine learning (ML) is a subset of data mining (DM). Data mining is the analysing of large data to the discovering of hidden trends, patterns and relationships, thereby developing a computer model to assist in decision making. Data mining is also termed as knowledge discovering in databases and datasets (Shun and Ryusuke, 2019) and it constitute a stage in Knowledge Discovery Database (KBB) process with other stages being data selection, data cleaning and evaluation. Manjunath (2015) referred to data mining as the core of knowledge discovering process. Machine learning is the ability for a computer or a program to learning from experience with respect to some tasks by the increase in the computer's or program's performance. A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

1.2 Problem Statement

Traditional anti-money laundering approach which employs the use of predefined scenarios (also called rules) to identify suspicious activities is irrelevant and obsolete in this era of big data. The processes of money laundering which are collection, placement, layering and integration; takes time, usually months or years never days nor weeks (Doppalapudi et al., 2022). This makes the detection and prevention of money laundering through the use of traditional approaches and explicit programming to be unbreakable, that is, difficult to detect and trace, and time consuming.

Crimes should be prevented and not allowed to happened. Most crime such as money laundering should not be allowed to happened, thereafter looking for the culprit after the crime has happened. Detecting money laundering only after the act has been done is a failure and problem on anti-money laundering system.

The frequency and volume of financial data for multinational business and conglomerate makes money laundering detection and prevention a big data issue, hence it is best solved using data mining and machine learning techniques and not traditional approaches and explicit programming.

Automated money laundering has been researched on using mostly machine learning methods (algorithms), rule base, semantics, neural network and fuzzy logic but these are mostly for detection purposes (Doppalapudi et al., 2022; Joana, 2015; Salehi et al., 2017). Hence, it becomes paramount to design a preventive model for money laundering.

1.3 Objectives

The aim of this study is to develop a web-based machine learning money laundering detection and prevention model.

The specific objectives are as follows:

1. To collect data through a financial dataset, having features like bank verification number (BVN) and national identification number (NIN) from Kaggle.com.
2. To analyse a financial dataset using python thereby removing any imbalance and unclean data.
3. To design a model using k-Nearest Neighbour (kNN) algorithm for detection of money laundering through Bank Verification Number (BVN) and Central Banks of Nigeria (CBN) bank account tiers limits.
4. To prevent money laundering by automating the anti-money laundering laws of the Central Banks of Nigeria (CBN).
5. To test and implement the design using metrics such as accuracy, precision, recall, f-measure and specificity.

1.4 Justification of Study

Money laundering detection systems or anti money laundering systems exist which helps to detect money laundering, but while wait till a crime is done before its detection? A preventive money laundering system or a preventive anti money laundering system is of importance.

At the conclusion of this research, the application of the findings and results will benefit Nigeria, other countries and the international communities, as they will be able to shutter or cover-up money laundering activities even before the crime take place.

Financial bodies or institutions that are saddled with the goal to prevent money laundering and to detect it, will find it easy to carry out this responsibility amidst the huge number of operations that occurs daily.

The Nigerian government for instance will be able to stop the activities of Internet criminal commonly called 'yahoo-yahoo', which is on the rise daily. The EFCC and ICPC will find their investigation and detection very easy since it will be done through the techniques of data mining – classification (kNN).

The aforementioned reasons hereby justify the necessity for this research work.

1.5 Scope of Study

The scope of this work is on prevention and detection of money laundering. This will be actualized by the implementation of kNN algorithms on a financial dataset by simulation. This research is limited to the features in the financial dataset since there is a limitation of banks not allowing connection to their customers' transaction dataset.

In this research, there is no attempt to prosecute money launders or to arrest criminals but only to investigate suspicious individuals and to report suspicious individuals. The arrest and prosecuting are left for the executive and judiciary arms of government.

CHAPTER TWO

LITERATURE REVIEW

A variety of approaches to the detection of money laundering have been done by different researchers. This will be reviewed in this chapter. Also, the concept surrounding machine learning will be discussed in this chapter. This chapter is organised into conceptual framework, empirical review, theoretical framework and summary of literatures reviewed.

2.1 Conceptual Framework

In this section, orientation to the study will be provided and align the elements of the study through the body of knowledge elucidated from existing literatures.

2.1.1 Money Laundering

Money laundering is a major problem for financial institutions, regulatory organisations, and law enforcement agencies. Money laundering has grown more relevant due to its involvement with a variety of illicit activities, including drug trafficking, terrorism financing, and tax evasion.

To a layman laundering simply mean cleaning. So, money laundering is simply the cleaning of money. This is, not to wash or dry money which is physically dirty, but to transform the nature of money which is psychologically dirty to clean money. A money is said to be dirty when it is obtained or got from an illegal activity, such as prostitution (in regions where it is not legal), Internet crime, destruction of national infrastructure, kidnapping, trafficking, gambling (in regions where it is not legal like India (Kharote and Kshirsagar, 2014)) and hacking. A ‘dirty’ money cannot be spent especially by those that claim to be responsible without laundering it, else it will call for questioning by the society.

The term ‘money laundering’ has been given different definitions by different authors. Notably, Alexandre and Balsa (2016) see it as a crime that typically turns certain illegal financial gains to legal gains. They went further saying that a set of financial and commercial operations characterize money laundering that is aimed at incorporating into economy illicitly derived goods, resources or values in a transitory or permanent way.

Kharote and Kshirsagar (2014) referred money laundering as washing of illegal (dirty) money through a cycle or iterative transaction to produce what appears to be a legal (clean) money. They stated that illegal money is that money which the bearer is not notified anywhere and that the rightful tax on that money is not paid. Another definition of money laundering is by Demetis (2018) which is, the masking of monetary gains resulting from any type of criminal activity. He stated that predicate offences are those criminal activities that are associated with money laundering, however, money laundering is a crime on its own.

When the source, nature, existence, location and disposition of money and/or property obtained illegally or from criminal activities such as embezzlement, drug trafficking, prostitution, advanced free fraud, corruption and large-scale crime; is concealed then money laundering has occurred (Ogbodo and Mieseigha, 2013). Zhiyuan et al. (2014) defined money laundering as the activities that disguise money received through illegal operations, therefore making them legitimate. Both disposition/movement of properties and acquisition/possession of any properties derived from illicit acts is money laundering (Ogbodo and Mieseigha, 2013).

The researchers mentioned defined money laundering economically but Ogbodo and Mieseigha (2013) added the justice point of view to the economic view. Whatever view the definition maybe,

money laundering is a criminal offence punishable by local and international laws and financial bodies. However, money laundering is a by-product of others crimes termed as associated crimes.

2.1.2 Money Laundering Process

1. Laundering of clothes with involves cleaning is done in a house. The house where money is laundered is a business, an investment or in a financial structure. Example of these are transportation companies, oil and gas, exchange houses, brokerage firms, trading organizations, car dealerships, casinos and so on.
2. Money laundering is a cycle or an iteration of transaction. The processes in the cycle of money laundering are collection, placement, layering and integration. This process is dynamic and it continuously mask the revenues from illegal activities gradually and over a long time.
3. Collection – after the successful execution of a crime, the criminal is paid. The payment for the crime might be ransom money for a kidnapped, the money from trafficking, drugs sales, bunkering and proceeds from other crimes. The collection is not limited to cash, as resources and goods can be collected but later transformed to cash. This process is usually done once.
4. Placement – the movement of the collected money by the collector to the place of investment, business, or a financial house. This is the first step in covering or disguising the act and the money launderer. Placement might involve more than once from a collection or once in different investments from a collection. This can be done internationally or locally. Smart money launders don't place their collection at once or they divide their collection and place it in different investments.

5. Layering – this is the other steps taken for deeper cover up and disguise (Frumerie, 2021). It becomes a continual of placement from the first placement but this time it is done by the 'house' where the first placement occurred and not by the launderer. This process makes the detection of illegal proceeds difficult for the law enforcement agencies.
6. Integration – over a said time, when the launderer feels that the money has been properly laundered, the 'clean' money is moved back into the economy either through purchase or through the bank from the laundering 'house' to the money launderer, that is the person that did the collection and placement. The integration process cannot be detected except with the help of an informant because the laundered money now appears to be profit from normal legal business.

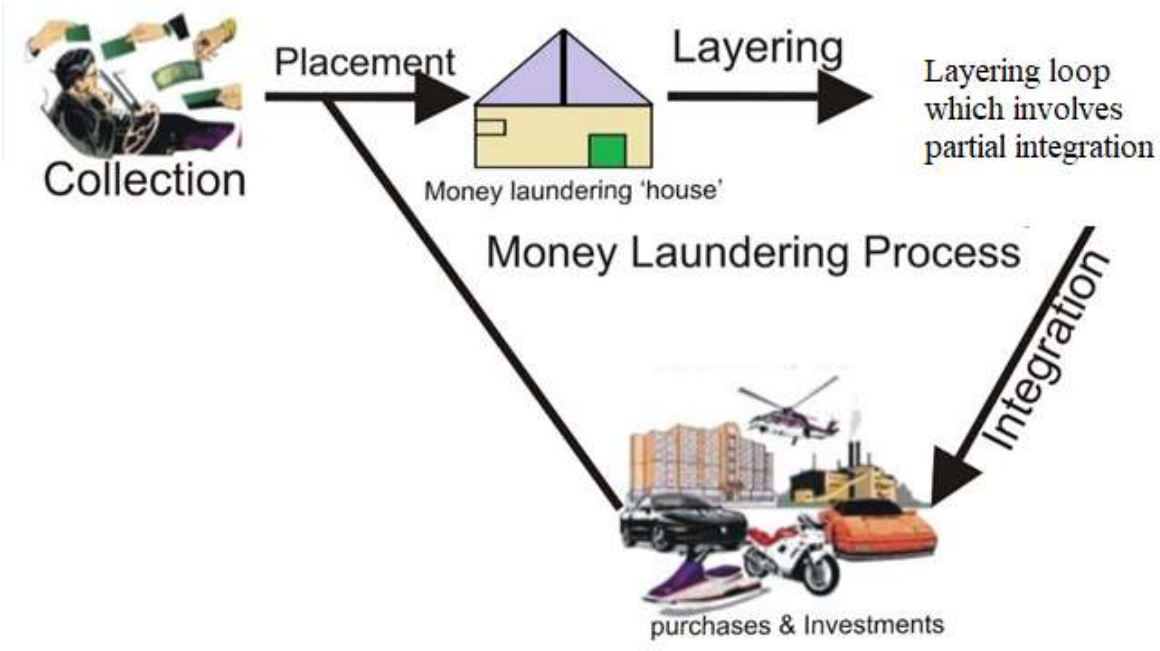


Figure 2.1 Money Laundering Processes (Manjunath, 2015)

Figure 2.1 illustrates the processes of money laundering which begins from collection. The money laundering house can be a legal investment, business or financial structure. The integration process is a loop and after the integration, the laundered money can be recollected by the owner or it can

be exchanged as purchases. One can insinuate that money laundering can be detected at the process of layering. For prevention of money laundering, process of placement is the point of prevention.

2.1.3 Anti-Money Laundering (AML)

The act of checkmating, detecting, discouraging, stopping and prosecuting money laundering and the money launderers through enacted laws and organized bodies is called anti-money laundering. Anti-money laundering can be a manual system or an automated system but it is all about a set of rules, procedures, laws and regulations to stop the practice of generating income through illegal actions (Doppalapudi et al., 2022; Manjunath, 2015).

When the anti-money laundering rules, laws and procedures are automated then it becomes anti-money laundering system. Cem (2023) and Manjunath (2015) defined AML system as a type of computer program that is used by financial institutions to analyse customer data and detect suspicious transactions. Furthermore, he stated that AML systems filter customer data, classify it according to level of suspicion and inspect it for anomalies.

Anomalies in an account can be obtained through the Know Your Customer (KYC) forms and other forms that were filled by the customers during the account creation or update. These anomalies can be an over limit deposit, a sudden increase in the balance of the individual, a large deposit into an account without legitimate trade that worth the deposited amount and others.

A smart money launderer makes his or her placement in bits or in different money laundering houses. This is done to avoid easy detection by the law enforcement agencies and by any automated system. Manjunath (2015) called the act of depositing large money in bit as structuring and we call the act of placement in different money laundering houses as dispersing.

Humans have been doing the work of detecting suspiciousness and anomalies in financial accounts of people before the advent of anti-money laundering systems. This they did and recorded little successes despite the small volume of data in those years (Cem, 2023). Businesses are on the increase daily and so are transactions, hence a human anti-money laundering system cannot monitor the transaction of such large companies since their transactions are in hundreds and at times thousands, and from different locations. If human anti-money launderers cannot successfully detect anomalies in large businesses, then what about conglomerates and multinational businesses with possibly thousands and tens of thousands of transactions daily? This definitely call for anti-money laundering systems that are programmed using data mining techniques since it is an era of big data.

2.1.4 Data Mining

1. In huge databases and data warehouses, there are different trends or patterns, and decision(s) has to be taken using these trends. These decisions that are needed to be taken might be for management use or for individual use or for research purposes. Some definitions of data mining are:
2. Data mining consists of finding interesting trends or patterns in large datasets, in order to guide decisions about future activities. (Banerjee et al., 2018)
3. Data mining refers to analysing massive amounts of data to uncover hidden trends, patterns, and relationships; to form computer models to simulate and explain the findings; and then to use such models to support business decision making. In other words, data mining focuses on the discovery and explanation stages of knowledge acquisition. (Jullum et al., 2020)

4. In a simple language, an automated process of locating and extracting the hidden patterns and knowledge in a large dataset is called data mining.

Data mining, also termed knowledge discovery in databases, is the procedure of learning stimulating and valuable forms and relations in huge dimensions of data (Salehi et al., 2017). Data mining is a stage or a process in the knowledge discovery database process (KDD). The other processes in KDD are data selection, data cleaning and evaluation; before evaluation is data mining (Figure 2.2).

a) Data Selection – Identify the target dataset and relevant attributes.
b) Data Cleaning – Remove noise, outliner, transform field values to common units, generate new fields through combination of existing fields, and bring the data into the relational schema that is used as input to the data mining activity
c) Data Mining – Extract the actual pattern
d) Evaluation – Present the patterns in an understandable form to the end user, for example through visualization.

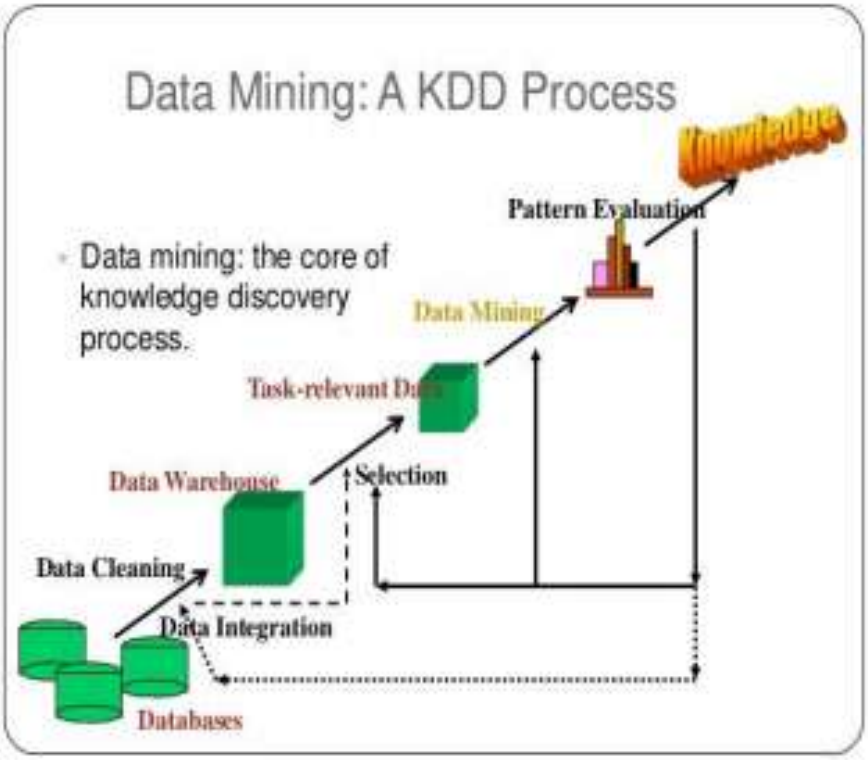


Figure 2.2 Knowledge Discovery Database Process (Manjunath, 2015)

Data mining can be run in two modes which are guided and automated. In the guided mode, the user makes use of the data mining tool step by step to explore and explain known patterns or relationships. The techniques to be applied to the data when in the guided mode is specified by the end user. In the automated mode, the user only set up the data mining tools which then runs automatically under covering hidden patterns, trends, and relationships. The data-mining tool applies multiple techniques to find significant relationships.

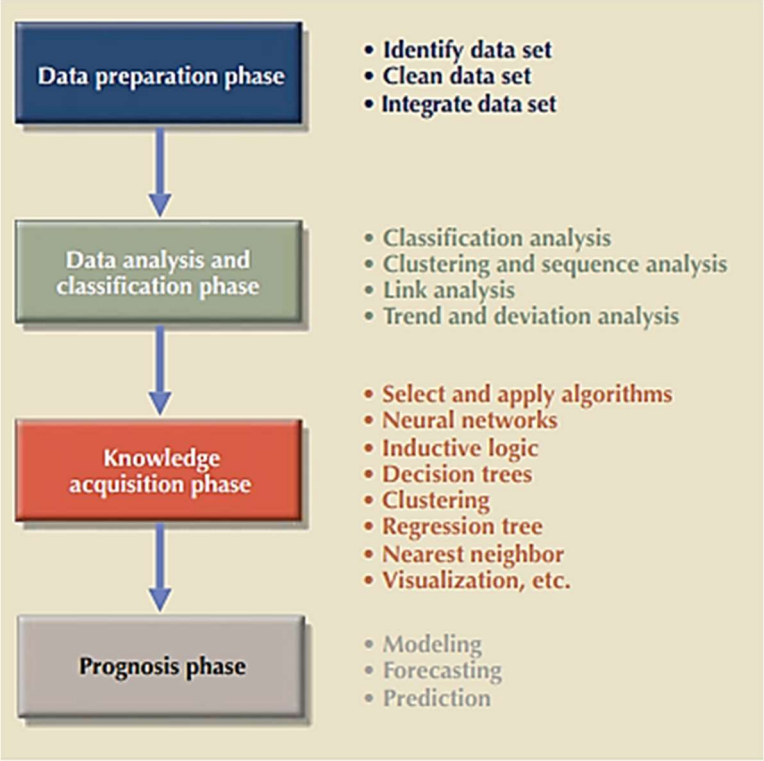


Figure 2.3 Phases of Data Mining (Carlos and Steven, 2017)

The phases of data mining as shown in Figure 2.3 are:

1. Data Preparation – Identify dataset, clean dataset and integrate dataset
2. Data Analysis and Classification – Classification analysis, cluster and sequence analysis, link analysis, trend and deviation analysis
3. Knowledge Acquisition – Select and apply algorithm, neural networks, inductive logic, decision trees, clustering, regression tree, nearest neighbour, visualisation etc
4. Prognosis – Modelling, forecasting, prediction

The primary goals of data mining in practice are prediction and description. Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest. Description focuses on finding human-interpretable patterns describing the data.

2.1.5 Machine Learning (ML)

Before the era of machine learning, systems were programmed explicitly hence their performance were exclusively on the written program. Machine learning being a branch of Artificial Intelligence (AI) is the practice of enabling a system to learn from data rather than through explicit programming. Machine learning came as a result of big data analytics. Predictive, analytic and descriptive models have to be improved therefore machine learning techniques are required (Ruiz and Angelis, 2022; Sujith et al., 2022).

Machine learning means to enable machines to learn without programming them explicitly. The objectives of machine learning are to enable machines make predictions, perform clustering, extract association rules, or make decisions from a given dataset.

According to I-Hsien et al. (2010), Lokanan (2022), Lopez-Rojas and Axelsson (2012) and Rafał et al. (2015), big data is any kind of data source that has at least one of four shared characteristics, called the four Vs:

a. Extremely large Volumes of data

b. The ability to move that data at a high Velocity of speed

c. An ever-expanding Variety of data sources

d. Veracity so that data sources truly represent truth

The types of learning in machine learning are as follows:

1. Supervised – It is concerned with labelled data. That is, it learns from a model of labelled (classified) data.

2. Unsupervised – It is concerned with unlabelled data. That is, it learns from a model of unlabelled (unclassified) data.

3. Semi-supervised – Here the machine learns from a mixture of both labelled and unlabelled data.

4. Reinforcement – Here the machine learns from no data.

A detailed look at supervised and unsupervised learning shows that in a supervised learning, the target is to infer a function or mapping from training data that is labelled. The training data consist of input vector X and output vector Y of labels or tags. A label or tag from vector Y is the *explanation* of its respective input example from input vector X. The X and Y make up the training example. In unsupervised learning, there is no training data and the idea is to find a hidden idea in the unlabeled data. In supervised learning, we have ML as either regression or classification while in unsupervised we have it as clustering (segmentation). Thus:

a) Regression task is to predict future value based on previous observation. Real or continuous numbers are regression task.

b) Classification task is based on set of label data. If the output is discrete or categorical then it is a classification task.

c) Clustering task is to find hidden pattern in unlabeled data and separates it into clusters according to similarity.

In machine learning, the algorithms can fall into any of the types of learning as shown in Figure 2.4. Table **2.**1 gives us a view of some machine learning algorithm under supervised and unsupervised learning

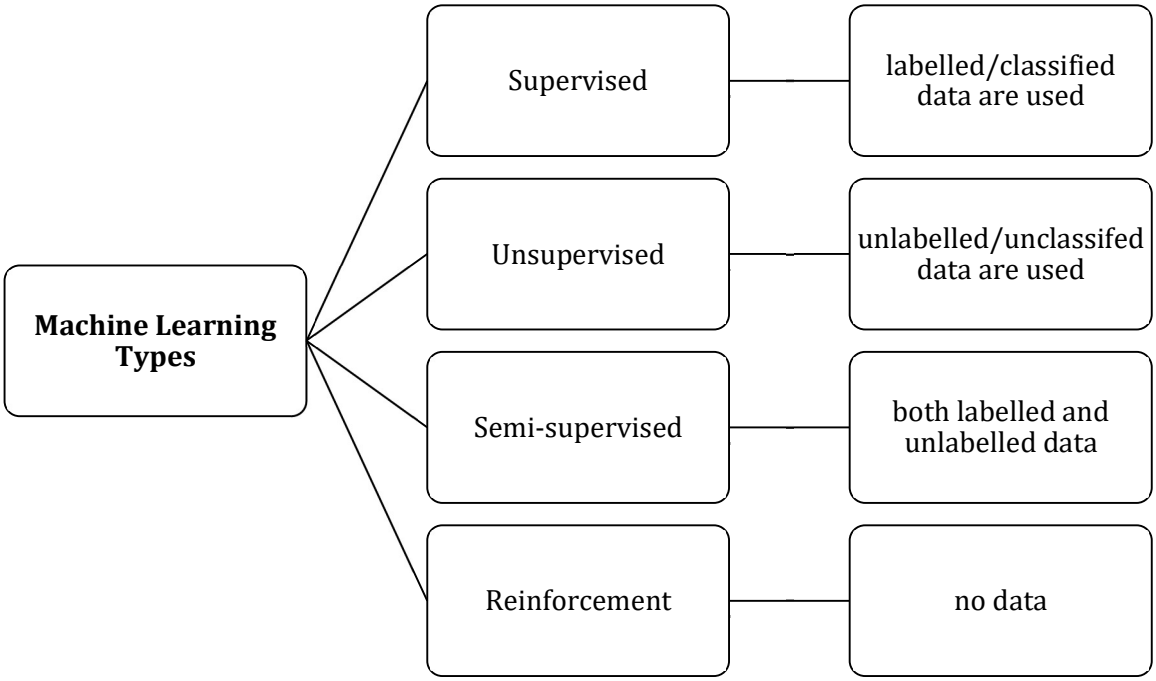


Figure 2.4 Machine Learning Types (Odii et al., 2019)

Table 2.1: Machine Learning Algorithms and Their Learning Type

S/N	Supervised	Unsupervised	Semi-supervised	Reinforcement
1.	Decision Tree	k-Means	Transductive SVM	Q-learning
2.	k-Nearest Neighbour	Hidden Markov Model	Manifold regularization	Deep Q Network
3.	Naïve Bayes	Principal Model Analysis	Laplacian regularization	Deep Deterministic Policy Gradient
4.	Artificial Neural Network	Gaussian Mixture Model	Laplacian SVM	Proximal Policy Optimization
5.	Linear Discriminant Analysis	Hierarchical clustering	Weight matrix	Asynchronous Advantage Actor-Critic
6.	Rule-Based Classifiers	Spectral clustering	Generative model	Trust Region Policy Optimization

Money laundering detection is based on supervised or unsupervised learning while money laundering prevention can be achieved using reinforcement or semi supervised learning.

The stages in machine learning are data intake, data transformation, model training, model testing and model development (Ahmed, 2019; Alarab et al., 2020; Llu´ıs et al., 2012) (see Figure 2.5).

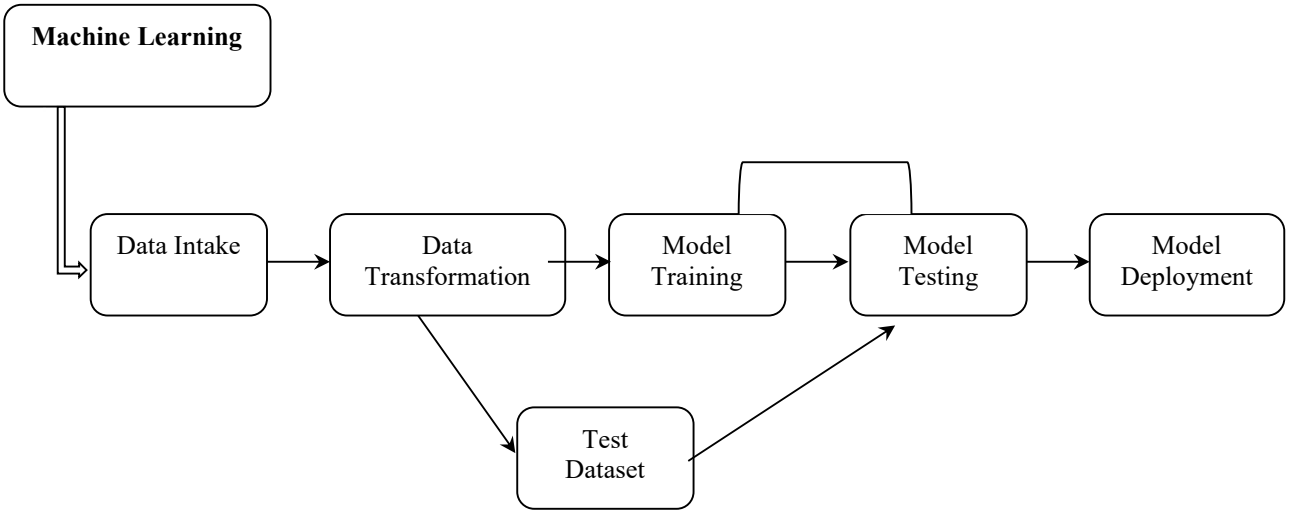


Figure 2.5 Machine Learning Stages (Odii et al., 2019)

i. Data Intake – A dataset is loaded and saved into the memory

ii. Data Transformation – Here the inputted data is transformed, cleaned and regularized. Data conversion to the needed format is done and the data is separate into training and test data. A training data is for building the model and testing data is for validating the built model.

iii. Model Training – A model is built using a chosen algorithm

iv. Model Testing – The built model is tested using test data and this is a continuous process because the produced result from the testing is used to build new model. This is the learning in machine learning.

v. Model Development – The best model after a desired iteration or one that fits the desired result is selected.

The input data which is mostly the dataset, have some attributes that must be extracted and fed into the machine learning algorithm. These attributes are referred to as FEATURES and the matrix is referred to as FEATURE VECTOR. The process of extracting the attributes (features) is referred to as FEATURE EXTRACTION. The features must be informative, non-redundant and relevant. After features are extracted, if the features are too big, the extracted features should be reduced to a feature vector with a smaller size. The process of reducing is referred to as FEATURE SELECTION. Extraction and selection are a process under data transformation which is the conversion of data that will be inputted into the algorithm to be suitable for the algorithm. Other data transformation processes are normalization, standardization and non-linear expansions.

Some algorithms used in machine learning are Naïve Bayes (NB), *k* Nearest Neighbour (KNN), Support Vector Machine (SVM), Hidden Markov Model (HMM), J48 Decision Tree, Random Forest and Artificial Neural Network (ANN).

1. Naïve Bayes (NB) – This relies on the Bayes theory of probability with strong independence assumption. It is used for classification task, both binary and multi-class classification. The features are treated independently. It is simple to implement and ease to understand but the downside is the treatment of the features independently (Kateryna, 2017). This method can be successfully used to the analysis of heterogeneous and highly dimensionality data, also, for spam mail detection.

2. k-Nearest Neighbour (kNN) – This is a sophisticated, simple and accurate ML algorithm. It is non-parametric that is, it makes no assumption about the data structure. It can be used

for both classification and regression task. It does its classification based on the principle of the nearest neighbours (samples). The downside is the bad performance on the unevenly distributed datasets (Kateryna, 2017). The classification is done in two phases; first the determination of the nearest neighbours and the second phase is to determine the class of a query sample based on the outcomes (assigned class). To classify an unlabelled object, the distance between this object and unlabelled object is computed and its k nearest neighbours are identified.

3. Support Vector Machine (SVM) – This is generally used for classification task. It finds a hyperplane that separates the classes in the best way. The distance between the 'support vector' (the points lying closest to the hyperplane) and the hyperplane is referred to as margin. SVM have a goal of finding a hyperplane that would result in the maximum margins. The upsides are good accuracy and its effectiveness in high-dimensional and noisy (overlapping) datasets but with larger datasets, training time can be high. SVM boils down to find a decision boundary – a plan (a hyperplane for $n > 3$), which divides data into two sets, one for each class. SVM can be extended to solve regression analysis, numerical calculations. It can also rank the elements and its insensitive to the outliers but the choice of the kernel can be the tedious task.

4. J48 Decision Tree – It have a tree like data structure and it can be used for classification and regression tasks. Iterative Dichotomiser 3 (ID3) is a common algorithm for decision tree. It is simple and handles large and noisy datasets well. It also operates in a white box. J48 makes it a popular solution for medical diagnosis, spam filtering, security screening and other fields.

5. Random Forest – It is very popular and requires almost no data preparation and modelling but usually results in accurate results (Kateryna, 2017). It is called a 'forest' because it is a set or collection of decision tress.

6. Hidden Markov Model (HMM) – These models are generally used for statistical pattern analysis though they can also be used in speech recognition, malicious code detection and biological sequence analysis. As a statistical model, it has states and known probabilities of the state transitions is called a Markov model. In a Markov model the states are visible but in a Hidden Markov model, the states are invisible. HMM is applied in detection cases notably the detection of metamorphic virus.

7. Artificial Neural Network (ANN) – This is a similitude of natural neurons hence an artificial neuron is a computational model inspired in the natural neurons. The goal of ANN is to process information. The ANN has been to detect Boot Sector virus using N-Gram as a feature.

2.1.6 Methods for Money Laundering Detection

Some methods of detecting money laundering are discussed in this subsection.

a) Machine Learning and Data Mining

Many recent research use machine learning and data mining approaches to examine financial transactions for possible money laundering (Zhang et al., 2021). Algorithms are used in these approaches to detect suspicious patterns, abnormalities, and unexpected behaviors that may be suggestive of money laundering.

The dependability of supervised machine learning methods in anti-money laundering (AML) applications was examined by Broussard and Wey (2020). They stressed how crucial AML initiatives are and how supervised learning may be used to identify questionable financial activity. The opacity and possible biases in these models, which may affect their reliability, are points of worry for the authors. The study emphasised how, in order to foster confidence and guarantee responsible and dependable execution in financial regulatory activities, AML machine learning models must be more transparent, interpretable, and equitable.

Li and Shi (2021) provided an in-depth examination of the application of machine learning techniques in the domain of anti-money laundering (AML). They investigate various aspects of AML, such as the challenges, techniques, and data sources relevant to the field. They present a comprehensive overview of machine learning methodologies used in AML, highlighting their strengths and limitations. Furthermore, the paper discusses emerging trends and future prospects in the use of machine learning. Vitas and Džemidžić (2019) explored the application of data mining methods to identify instances of money laundering. The study looks into how data analytics might be used to spot questionable activities and patterns in financial systems. They addressed the drawbacks of using conventional techniques to identify money laundering and emphasizes data mining's potential as a cutting-edge, pre-emptive strategy. The usefulness of data mining techniques, including clustering, classification, and anomaly detection, in identifying anomalous financial activities that could be signs of money laundering is investigated. They also took into account how crucial feature selection and data quality are to the effectiveness of data mining applications in anti-money laundering campaigns. To increase the detection system's accuracy and efficiency, they stress the necessity of high-quality data and the identification of pertinent characteristics.

Khan and Verma (2021) emphasized the growing importance of big data analytics in the fight against money laundering, especially in light of the massive volumes of transaction data that authorities and financial institutions need to handle. Big data analytics makes it possible to spot unusual behaviour, intricate patterns, and questionable activity that can point to money laundering. The authors went over the several uses of big data analytics in AML, such as risk assessment, client profiling, and transaction monitoring. They also took into account the difficulties and constraints that came with using big data analytics in AML, including issues with data quality, privacy, and the requirement for sophisticated analytical skills and tools. The significance of artificial intelligence and machine learning methods in relation to big data analytics and AML. They offered a thorough rundown of these technologies' possible advantages as well as how they might improve AML procedures.

b) Blockchain Analysis

Researchers are looking at blockchain technology in an effort to track and identify instances of money laundering involving digital currencies due to the surge in popularity of cryptocurrencies (Liu et al., 2020).

Shcherbakov and Smith (2018) investigated the core ideas of blockchain and digital currencies, as well as the unique dangers and hazards connected with money laundering in this new financial landscape. According to the report, the anonymous and decentralized character of blockchain transactions made them appealing for illegal financial activity. The paper addressed numerous money laundering strategies inside the cryptocurrency ecosystem, such as tumbling and mixing services, as well as the difficulties that law enforcement authorities have in locating and apprehending money launderers. It also looked at legislative measures and international attempts

to fight money laundering in the blockchain and cryptocurrency area. The authors ended by providing insights into current developments and the need for strengthened regulatory frameworks to combat the rising money laundering dangers in this arena.

Kamp and Klein (2020) analysed the rising risk of money laundering in the domain of cryptocurrency. The authors investigated how cryptocurrencies' distinctive qualities, such as anonymity and decentralized transactions, render them vulnerable to money laundering operations. The article explored numerous money laundering strategies used within the cryptocurrency ecosystem, such as mixing services and privacy coins, and emphasised the rising usage of digital currencies for criminal financial goals. It also investigated the difficulties that regulators and law enforcement organizations confront in tackling this increasing threat, highlighting the importance of enhanced detection and preventive methods. To address the developing picture of cryptocurrency-based money laundering, the authors concluded by emphasising the significance of international collaboration and regulatory frameworks.

c) Network Analysis

In order to detect suspect financial flows, some research concentrates on network-based techniques that take into account the links and linkages between people, things, and transactions (Elham et al., 2019).

2.1.7 Regulatory and Compliance Measure

Some regulations to check and curtain money laundering are presented in the following points.

a) Anti-Money Laundering (AML) Regulations

To improve the identification and reporting of money laundering, regulatory agencies and financial institutions regularly revise anti-money laundering (AML) legislation (Mayhew et al., 2021). In Nigeria, Independent Corrupt Practices and related Offences Commission (ICPC) and Economic and Financial Crimes Commission (EFCC) alongside the Central Bank of Nigeria are anti-money laundering regulatory bodies.

Brust (2019) looked at the changing environment of anti-money laundering (AML) procedures. The report highlighted and explored six significant compliance developments that arose in 2019. Changes in regulatory requirements, the use of modern technology, and a more proactive attitude to AML initiatives are all examples of these developments. The report emphasised the significance of adopting digital AML solutions, such as the use of artificial intelligence and machine learning for risk assessment and monitoring. It also investigates the growing emphasis on international collaboration in AML activities, as well as the influence of expanding regulatory requirements on financial institutions. The author closes by underlining the need of firms adapting to these changes in order to improve their AML compliance operations.

Sahadev and Gupta (2021) conducted a comprehensive literature study to investigate the relationship between anti-money laundering (AML) compliance and technology. The research examined and synthesised major insights from previous studies on the use of technology in AML initiatives. It emphasised the growing relevance of technical tools in upgrading AML procedures, such as artificial intelligence, machine learning, and blockchain. They further identified a number of themes and trends in the literature, such as the difficulties of maintaining regulatory compliance in the face of quickly advancing technology, the automation of AML compliance duties, and the

efficacy of technology-driven risk assessment. The study also proposed a research agenda to direct future investigations in this area, highlighting the necessity of conducting more thorough and comparative evaluations of AML technology solutions and their effects on the financial sector.

b) Technological Innovations in Compliance

The significance of technology, including RegTech and AI-driven compliance solutions, in improving money laundering detection and guaranteeing regulatory compliance is also covered in recent research (Braun et al., 2018). RegTech which stands for Regulatory Technology is a new technology that entails putting digital tools and procedures into place to help businesses better manage their growing regulatory compliance obligations.

Stojanović and Milovanović (2020) investigated the impact of financial technology (FinTech) on the regulatory environment and money laundering operations. Peer-to-peer lending, blockchain technology, digital payment systems, and other FinTech technologies have changed the financial environment, and this paper examines how this has happened. Financial institutions and regulators now face both possibilities and problems as a result of these changes. The study explores the ways in which money launderers could use FinTech innovations to hide their illegal activity. It also looked at how regulatory bodies adjusted to these developments and improved anti-money laundering (AML) procedures. In the context of FinTech, the writers talked about how crucial it is to take a proactive approach to AML, utilising blockchain, artificial intelligence, and data analytics to improve compliance.

The study also assessed how well regulatory initiatives are working to keep up with the rapidly evolving FinTech landscape, highlighting the necessity of frameworks that balance security and innovation. In their last section, the writers offered their perspectives on the current issues and

potential solutions related to money laundering and fintech. Colombo and Aicardi (2021) conducted a thorough scoping analysis to assess the volume and kind of studies on the application of financial technology (FinTech) to money laundering prevention. The study's objective was to present an overview of the body of research in this area, highlighting important trends and gaps. They examined a variety of studies, publications, and research articles about the role that fintech plays in preventing money laundering. They emphasised how important FinTech innovations—like blockchain, artificial intelligence, and data analytics—are to strengthening anti-money laundering (AML) procedures and bolstering the capacities of regulators and financial institutions. They also discussed the several ways FinTech is used in AML, including as transaction tracking, customer due diligence, and identifying questionable financial activity. It also looked at the difficulties and restrictions that come with using FinTech solutions in AML and the requirement for suitable legal frameworks to control these technologies. The scoping assessment concludes by highlighting the possible advantages and challenges of integrating FinTech into AML procedures and offering insights into the changing FinTech ecosystem and its role in combating money laundering.

Ghosh and Chaudhary (2020) presented an empirical investigation of the connection between money laundering and digital identities. They investigated how the use of digital identity verification techniques affects the fight against financial crime and money laundering. This study looked at how well digital identification solutions work to improve customer due diligence (CDD) procedures and lower the chance of money laundering. In order to increase the precision and effectiveness of identity verification, it evaluated the benefits and drawbacks of digital identification technologies, such as biometrics and digital authentication. They also investigated the issues and weaknesses that might surface while using digital identity verification, especially in

relation to privacy and data security. It examined how financial institutions and regulatory agencies might handle these concerns as well as the possible exploitation of digital identification information in money laundering operations. In conclusion, the study offered insightful information on the function of digital identities in the battle against money laundering. It also offered suggestions for putting in place efficient digital identity verification procedures together with empirical support.

Kabir and AlJamea (2021) examined the application of biometrics in attempts to combat money laundering (AML). They seek to give a summary of the literature already in existence in this field, point out knowledge gaps, and suggest possible directions for further research. They highlighted the increasing significance of biometric technology, including face recognition, fingerprint recognition, and iris scanning, in fortifying anti-money laundering procedures. Using biometrics can help reduce the risk of fraud and money laundering by securely and reliably confirming the identification of those who are involved in financial transactions. The article's authors go over a number of biometric uses in AML, including as better due diligence, transaction monitoring, and client onboarding. They also took into account the difficulties and moral issues related to the gathering, storing, and use of biometric data. In order to further their paper, the authors suggested a research agenda that would direct future investigations into the intersection of biometrics and AML. They emphasized the importance of empirical research, the assessment of the efficacy of biometric systems, and the creation of best practices for the integration of biometric solutions into AML procedures. The study concluded by offering insights into the potential of biometrics as an AML tool and outlining future research priorities in this field.

2.1.8 Regulatory and Compliance Measure – Issues

Two vital issues with the compliance of anti-money laundering regulations are stated in this subsection.

a) Privacy and Ethical Concerns

The problems of gathering and evaluating personal financial data are examined in recent research, and money laundering detection attempts must strike a compromise between privacy and ethical issues (Nikolov et al., 2022).

b) Global Cooperation

Recent study emphasizes the necessity for more international collaboration and information sharing among law enforcement organizations since money laundering frequently crosses international borders (Shaw et al., 2023).

2.1.9 Financial Transaction in Nigeria

The Central Bank of Nigeria (CBN) regulates the Deposit Money Banks (DMB) and other financial institutions. Laws concerning financial transactions are made by the CBN such as the Anti-Money Laundering/Combating the Financing of Terrorism (AML/CFT) Policy and Procedure as amended in February, 2019. The AML/CFT law defined money laundering as the act of directly or indirectly concealing or disguising any fund or property that is derived from the proceeds of an unlawful activity. Simply put, it is the process by which “dirty” money is made to look legitimate or “clean” so that funds may be used freely without any trace of its illicit source (CBN, 2019).

Financial transactions as stated in AML/CFT Act 2019, must be conducted through correspondent banking relationships in conjunction with a risk-based approach, and Know Your Correspondent (KYC) procedures (CBN, 2021a). Despite all financial transactions following one channel, it is the duty of the DMBs to keep records of the financial records (CBN, 2021b). Financial Transactions can be done in-bank and online using channels like Unstructured Supplementary Service Data (USSD) code, e-banking/online banking, mobile app, Nigeria Inter-Bank Settlement System (NIBSS) Instant Payment (NIP) and NIBSS Automated Payment Services (NAPS).

Account types and their limits differ from one Deposit Money Bank (DMB) to another; however, they are not in contrast to the three (3) standard account types in Nigeria as ordered by the CBN (CBN, 2021c). The account types (tiers) in Nigeria as stipulated by the CBN, the KYC procedure and their conditions are shown in Table 2.2.

Table 2.2 CBN Account Tiers, KYC Procedure and Conditions (Kuda, 2021)

S/N	Account	KYC	Condition
1.	Tier 1 (Low Value Account)	Basic Information and may not be verified by DMB	● ₦ 20,000 maximum deposit ● ₦ 200,000 maximum balance ● ₦ 3,000 online single transaction ● ₦ 30,000 online maximum daily limit ● No transfer ● No International operations ● Strictly Savings
2.	Tier 2 (Medium Value Account)	Basic Information and must be verified by DMB	● ₦ 50,000 maximum deposit ● ₦ 400,000 maximum balance ● ₦ 10,000 online single transaction ● ₦ 100,000 online maximum daily limit ● Transfer only within Nigeria ● No International operations ● Strictly Savings
3.	Teir 3 (High Value Account)	Full KYC requirement (National ID or Voter's Card or International Passport, business registration document and Utility Bill which Address must be verified by DMB) and other basic information	● No maximum deposit ● No maximum balance ● ₦ 100,000 online single transaction ● ₦ 1,000,000 online maximum daily limit ● Transfer to any possible country ● International operations ● Savings, Current or Corporate Account

The CBN through its laws that are transformational seasonally (CBN, 2021d), check mates the acts of money launderers. However, despite the laws in different financial domain, there are still launderers of money, that is why the CBN periodically institutionalize laws against money

laundering (CBN, 2021e), though money laundering is given different names in Nigeria such as embezzlement, conversion and misappropriation of funds (NBS, 2016, 2017).

2.1.10 Money Laundering Cases in Nigeria

Money laundering in Nigeria is prominent especially among the politicians, as can be seen in the records of the Economic and Financial Crimes Commission (EFCC) in Table 2.3. The commission between 2002 and 2021 reported more than 975 cases related to financial crimes of which 43 are high profile cases and 16 are high profile money laundering cases (Ahmed, 2019; EFCC, 2019, 2021). Some high-profile cases recorded in recent times are EFCC V DR. MARTINS OLUWAFEMI THOMAS in 2018, OYEBODE ALADE ATOYEBI V. FEDERAL REPUBLIC OF NIGERIA in 2017, Ude Jones Udeogu V. Federal Republic of Nigeria and Others in 2016.

Table 2.3: Some EFCC Crime Report

S/N	Detail	Amount	Date	Source
1.	Ismail Mustapha Vs EFCC	N33 billion	November 25, 2019	https://guardian.ng/news/court-fixes-jan-15-to-try-mompha-for-alleged-money-laundering/
2.	3 Bankers Vs Lagos State Police	US$450,267.82 (N162,765,000)	23 January 2020	https://guardian.ng/news/police-quiz-bank-workers-for-alleged-fraud-money-laundering/
3.	Mohammed Adoke Vs EFCC	Undisclosed	10 February 2020	https://guardian.ng/news/alleged-money-laundering-again-efcc-arraigns-adoke/
4.	Jumoke Akinjide and Others Vs EFCC	N650 million	January 16, 2018	https://guardian.ng/news/trial-of-ex-fct-minister-others-to-resume-october-11/
5.	Ayodele Fayose Vs EFCC	N1.2bn	October 22, 2018	https://guardian.ng/news/alleged-n1-2bn-laundering-efcc-presents-11th-witness-against-fayose/
6.	Dr Abdu Bulama and four others Vs EFCC	N229 million	15 October 2018	https://guardian.ng/news/ex-minister-4-others-reappear-in-court-over-alleged-n229m-money-laundering/
7.	Yemi Akinwonmi, Dickson Atiba and Ogunmodede Oladayo Vs EFCC	N179.8 million	March 2015	https://guardian.ng/news/efcc-arraigns-three-inec-officials-for-allegedly-laundering-n180-million/

2.2 Theoretical Framework

Money laundering being a world epidemic needs to be addressed. There are different models that see to address the issues of money laundering. Joana, (2015) gave a risk assessment model for anti-money laundering system. The aforesaid model is a client behaviour model based on clusters, also using Prospero software, which with Random Forest and the method of Self Organizing Maps allows to group clients according to their behaviour in terms of transactions. Her model allows each cluster to have an associated risk, and those with higher risk are targets of special attention. Whenever there is a deviation from the expected behaviour, there is a warning of suspected money laundering. The model is shown in Figure 2.6.

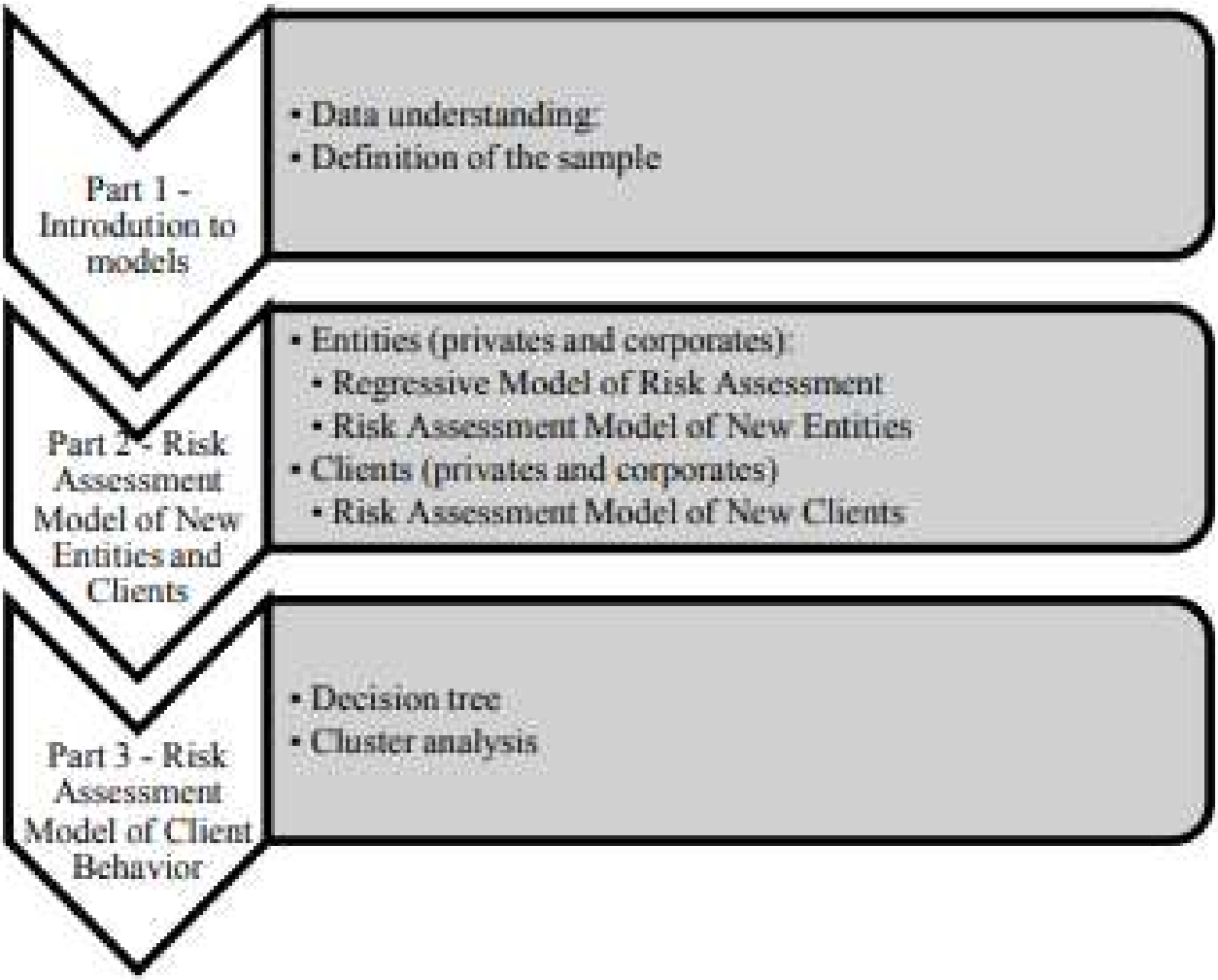


Figure 2.6 Risk Assessment Model (Joana, 2015)

In a knowledge discovery database (KDD) process, anti-money laundering is a process of four types and these steps corresponds to the four levels of analysis: transaction, account, institution and multi-institution (Le-Nhien et al., 2010). The first three levels: transaction, account and institution are the most important where the last one depends more or less on the organisations and their policy. The data mining framework proposed by then also consist of three levels and consist of different components. The three levels are: data pre-processing, data mining and knowledge management.

Data Pre-Processing

The main role of this component is to extract and clean raw datasets from data sources located in different sites of this international bank. It then integrates them into consolidated databases that are used to build a data warehouse of customer information and customer transactions.

Data Mining

This component provides classification and clustering techniques for the most basic level of this framework: analysing transaction datasets. At this level, transaction records are extracted for investigations.

Knowledge Management

Results of mining process, experience of AML experts, running results are collected, stored in relevant repositories and analysed by this component. It also generates significant, interpretable rules and knowledge.

Le-Nhien et al. (2010) also gave a diagrammatic representation for their process. They applied a clustering technique for the analysing and investigating process of the anti-money laundering system. This is shown in the Figure 2.7:

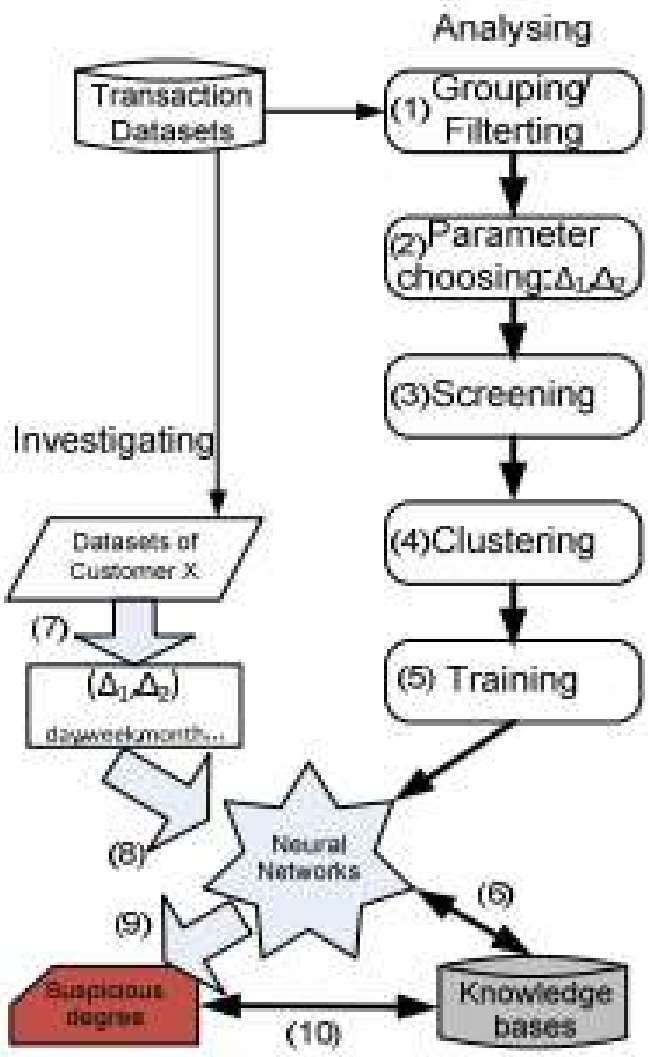


Figure 2.7 Analysing and Investigating Process (Le-Nhien et al., 2010)

Another model as studied by Timm et al. (2016) is the Customer Identification Process in Figure 2.8. The CIP model is like the Know Your Customer (KYC) model. The CIP is triggered every time the institute enters a new business relationship with a customer just like the KYC.

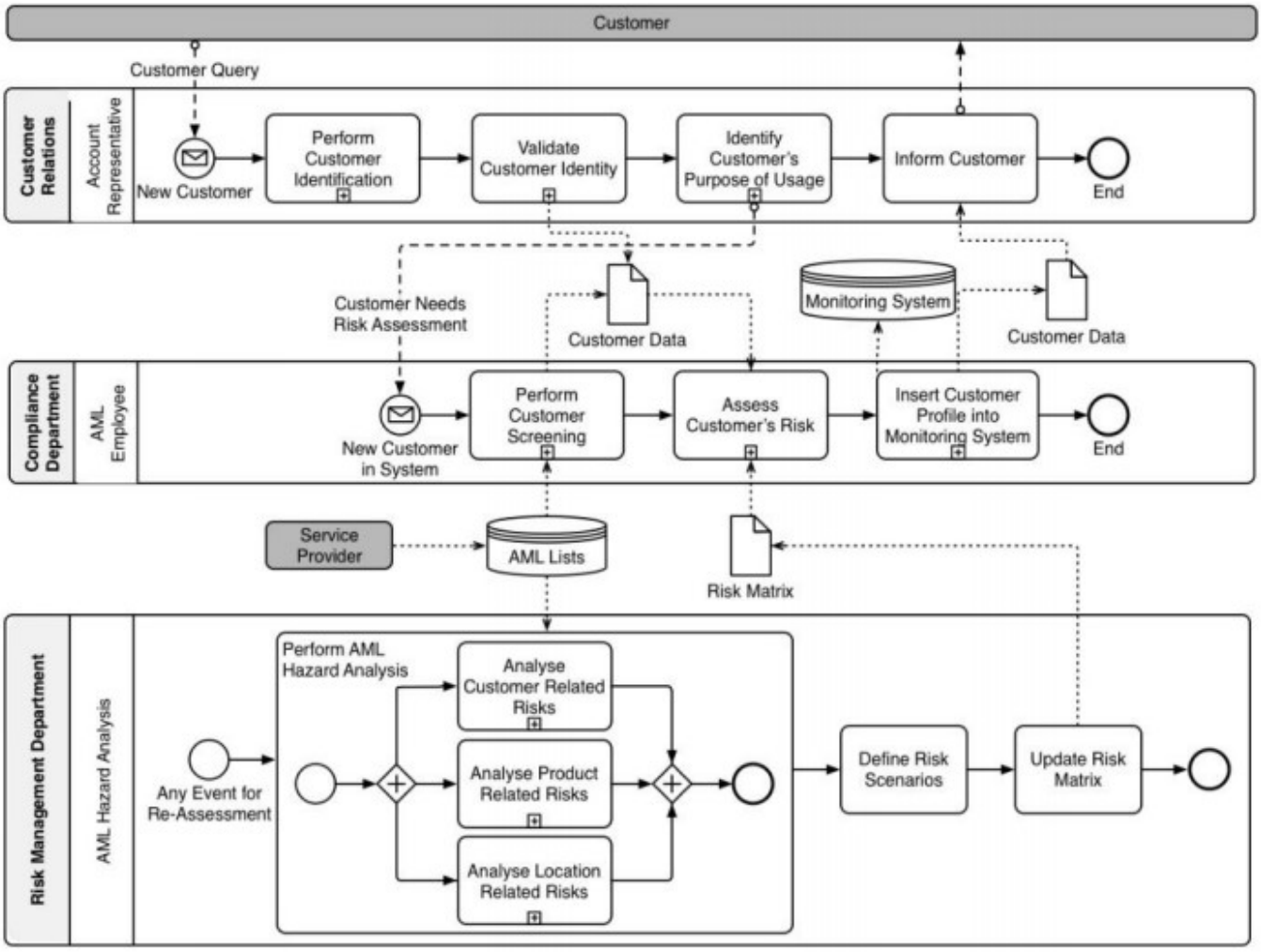


Figure 2.8 CIP in An AML (Timm et al., 2016)

The development of a reference model is an iterative process. This process is characterized by different versions of the considered model. The reference model should be evaluated using a validation method, which may lead to adjustments of the reference mode. Timm et al. (2016) used two iteration loops were traversed. While the first iteration loop concentrated on the process perspective, the second iteration loop focused on the data perspective of the AML program.

2.3 Empirical Framework

Money laundering as a world financial epidemic give both researchers and non-researcher a great concern. Many authors used different approaches to extensively research on detecting money laundering but the problem of money laundering still persist.

Xingqi and Guang (2009) proposed a system based on improved minimum spanning tree clustering to detect money laundering. According to them their system (algorithm) is effective and succinct but their system only detected without preventing. Their system used a financial dataset which have 70 fields, 64941 records and it was coded in Microsoft Visual C++. Their work was an analysis of similarity measure and distance metric.

Intelligent agent-assisted decision support system was modelled and developed by Shijia and Dongming (2009) for anti-money laundering system. They decided to use intelligent agents because they are autonomous, reactive, proactive and they are suitable for dynamic, ill-structured and complex money laundering. They were unable to perform task analysis and knowledge acquisition in their system.

Le-Nhien et al. (2010) in their research applied a data-mining based solution to detect suspicious money laundering cases in an investment bank. They stated that traditional approaches to anti-money laundering followed a labour-intensive manual approach because money laundering is a sophisticated activity with many ways of laundering money. Hence, they proposed an investigating process to money laundering by using the clustering and neural networks algorithm of data mining. In addition to the two proposed algorithms, there was need to improve the running time of their system, hence they introduced suspicious screening heuristics.

Synthetic data approach was applied by Lopez-Rojas and Axelsson (2012) for the detection of money laundering. However, their approach was faced by the cons of synthetic data which include; biasness, non-realistic, non-representativeness of the data. Their system was a mobile system.

In the prediction model or technique, Lopez-Rojas and Axelsson (2012), analysed the implications of using machine learning techniques for money laundering detection in a data set consisting of synthetic financial transactions and aimed to detect anomalies inside a data set of mobile money financial transactions by using the classification techniques to group transactions as suspicious or nonsuspicious.

Zhiyuan et al. (2014) explored the benefits and effectiveness of expectation maximisation algorithm for suspicious transaction detection in anti-money laundering and they discovered that EM method out phased the traditional clustering method. They showed that EM (Expectation Maximisation) is a better algorithm to support clustering operation than k-means. However, their system was based on clustering alone, and it has a high false positive rate. Hence some normal transactions were seen as abnormal.

Clustering, classification and prediction were the data mining techniques identified by Manjunath (2015) that can be used by banks for the purpose of uncovering money laundering trends in their large financial database.

Alexandre and Balsa (2016) in their research used data mining techniques. This was done to support the process of detecting money laundering operations in a bank. They employed the WEKA tool (Waikato Environment for Knowledge Analysis) which is a product of the University of Waikato, New Zealand. The WEKA tool uses k-means, J48 and PART algorithms.

Timm et al. (2016) in their research classified the anti-money laundering programme for financial institutions into planning and controlling. The steps in Table 2.4 should be used in developing an anti-money laundering system.

Table 2.4 AML Program for Financial Institution

Phase	Step	Name
Planning	1.	Identify regulations
	2.	Derive company guideline
	3.	Conduct risk analysis
	4.	Define process and control activities
	5.	Implement control system
	6.	Define control structure
Controlling	7.	Define organization function
	8.	Appoint representative
	9.	Conduct employee training
	10.	Conduct internal and external audits

A graph-based machine learning approached was implemented by Soltani et al. (2016) in their money laundering detection framework. Their aim was to mine a cluster or group of transaction that have the features of money laundering with respect to their dataset and their rules. After the execution of their framework, a human investigator has to perform further analysis on the groups obtained by the framework.

Demetis (2018) reported the situation of banks trying to combat money laundering through structural coupling and machine learning profiling. In his work, the critical dynamics between computer profiling and human profiling was presented.

Another method for money laundering detection was applied by Frumerie (2021). This method entails the use of a boosting algorithm called XGBoost and a graph algorithm called graph convolutional networks. The experiment was conducted on two simulated financial datasets (carefully constructed synthetically generated datasets – PaySim and AMLSim) but with similitude as real financial data, and it shows that the XGBoost performs better than the graph convolutional networks (GCN), having an accuracy of 99.7% on PaySim and 99.9% on AMLSim datasets respectively.

A framework was proposed to detect fraudulent activities via the use of machine learning algorithms (Ramya et al., 2022). The framework explains the collaboration between banks by the increase in the authorities that money laundering detection data, in order to achieve higher success rate. In the absence of collaboration between banks, a fraudulent practitioner operates in an independent and isolated environment, thereby giving rise to increase in the risk of money laundering. Also, they suggested a central database for financial transactions.

Machine learning and deep learning was used by Alotibi et al. (2022). The algorithms used were Deep Neural Network (DNN), random forest (RF), K-Nearest Neighbours (KNN), and Naive Bayes (NB) on a bitcoin elliptic dataset. Their research showed that the highest performance was recorded in the DNN and RF models, however, the RF outperformed the DNN model.

2.4 Summary of Literatures Reviewed

A review has been done on money laundering, anti-money laundering and machine learning as researched by different authors.

Table 2.5 Summary of literatures reviewed

Author(s) and Year	Title	Work done	Limitation
Alotibi et al., (2022)	Money Laundering Detection using Machine Learning and Deep Learning	Employed Deep Neural Network (DNN), random forest (RF), K-Nearest Neighbors (KNN), and Naive Bayes (NB) on a bitcoin elliptic dataset to build AML model	Imbalanced data Data is not properly labelled
Ramya et al. (2022)	Comparative Analysis and Implementation of AI Algorithms for Money Laundering Detection	Designed a framework for the detection of money laundering using a dependent environment	Work was not implemented.
Frumerie (2021)	Money Laundering Detection using Tree Boosting and Graph Learning Algorithms.	Used XGBoost and convolutional network to detect money laundering on two simulated financial datasets	Large data is required due to the convolutional network.
Demetis (2018)	Fighting money laundering with technology: A case study of Bank X in the UK	Reported the situation of banks fighting money laundering through structural coupling and machine learning profiling	There is a critical dynamics between computer profiling and human profiling as presented
Soltani et al. (2016)	A New Algorithm for Money Laundering Detection Based on Structural Similarity	Used graph-based machine learning	Prevention not enabled and human effort still needed

Timm et al. (2016)	Building a Reference Model for Anti-Money Laundering in the Financial Sector	Used rule-based approach for the detection of money laundering	System cannot self learn and discover hidden patterns
Alexandre and Balsa (2016)	Client Profiling for an Anti-Money Laundering System	Used WEKA tool to profile bank customers in order to detect money laundering activities	There system was not for prevention purposes
Manjunath (2015)	Data Mining Techniques for Anti Money Laundering	Identified clustering, classification and prediction as data mining techniques for money laundering detection	Didn't no practical work.
Zhiyuan et al. (2014)	Exploration of the Effectiveness of Expectation Maximization Algorithm for Suspicious Transaction Detection in Anti-Money Laundering	Detected money laundering activities by the application of EM algorithm which is better for clustering	High false positive rate
Lopze-Rojas and Axelsson (2012)	Money Laundering Detection using Synthetic Data	Applied synthetic data for money laundering detection	biasness, non-realistic, non-representativeness of the data
Le Nhien et al (2010)	A data mining-based solution for detecting suspicious money laundering cases in an	Applied data mining (clustering and neural networks) to detect suspicious money laundering in a investment bank	Need to improve running time.

	investment bank		
Shijia and Dongming (2009)	Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering	Used intelligent agents for money laundry detection	System was unable to perform analysis and knowledge acquisition
Xingqi and Guang (2009)	Research on Money Laundering Detection Based on Improved Minimum Spanning Tree Clustering and Its Application	Used improved minimum spanning tree clustering to detect money laundering	Didn't prevent money laundering activities. It measures distance metrics only

Rule based systems, data mining and machine learning were used to address the issue of money laundering detection, however, the idea of money laundering prevention was not addressed by the authors of the reviewed literatures. The identified research gap is in the prevention of money laundering via machine learning techniques.

2.5 Research Gap

The reviewed and existing researches applied different techniques to detect money laundering, but the prevention of money laundering was left unattended. The identified research gap is the inability of different researches to prevent money laundering. Hence, this research will fill this gap by developing a web-based model that can be used to prevent money laundering and as well as detect money laundering too.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Software Methodology

Rapid Application Development (RAD) software methodology was applied in the design of the web-based model for money laundering detection and prevention. RAD is a software development approach; speed of releasing prototypes and modules is prioritized. User's feedback and responses is relied on in RAD for each released feature at the end of every iteration and the solution is pushed back into the pipeline immediately.

RAD can be broken down into the following steps

1. Define user requirements

   The user requirement is to detect money laundering and also to prevent same from financial transactions.

2. Developing prototypes

   Different prototypes were developed to fit in the newly engineered features and processing from the model. The different prototypes were the different hyperparameter (such as the value for metric and n_neighbours) applied in the kNN algorithm experimented during the model building.

3. Feedback collection

   Using a questionnaire (see Appendix E), feedback was collected. The collected feedbacks were used to structure the programming of the preventive model.

4. Testing

Testing of the model was done using 15% of the observations in the datasets.

5. Deployment

   The model was deployed using Flask framework on the local host and also on web.

3.2 Research Instrument

Questionnaires were administered for the purpose of data collection with respect to the laws and conditions for money laundering in Nigeria. The questionnaires (see Appendix E) were administered online and offline. Respondents to the questionnaires was specified to be any of the following roles – customer service branch manager, bank branch manager, fraud control unit branch manager. Forty-five (45) email addresses of different financial institutions were sent the Google form questionnaire. Five banks were contact in person (physically). The contacted banks are Fidelity bank, Guaranty Trust Bank, Ecobank, Zenith Bank and Unity Bank The questionnaire was a mixed question, consisting of open ended and closed ended (Yes, No) questions.

3.3 System Analysis

The existing system and the proposed system were analysed in the subsections.

i. Analysis of the Existing System

The existing system is a model built using KNN adopted by Alotibi et al. (2022). They employed k-Nearest Neighbour (kNN) algorithm in building their model which gave them a f1-score of 97%, accuracy of 92%, precision of 97% and a recall of 98%. The existing system was trained using the dataset in the ratio of 70% for train set and 30% for the test set.

ii. Weakness of the Existing System

The existing system was trained with a highly biased (imbalanced) data; 157,205 samples of the data were not labelled while 4,545 were labelled as illegal transaction and 42,019 were labelled legal transactions. Their final selected features were 53 and this is too large thereby leading to overfitting. Hence, the fairness of the system is questionable as it will learn more of legal transaction of the legal samples since the authors did not balance the dataset. The authors only standardized after applying correlation but they did not balance the dataset's label.

iii. Analysis of the Proposed System

Scikit-learn is an open-source python machine learning library. It features several algorithms like the linear regression, support vector machine, random forest, k-means and k-nearest neighbours; and it also supports Python numerical, visualization and scientific libraries like Pandas, Matplotlib and NumPy. k-Nearest Neighbour (kNN) as a machine learning model that was used in building the model consist of storing only the training dataset, and prediction for a new data point is made by finding the closest data points in the training dataset. The closest data points are referred to as the nearest neighbours to the new data point.

The dataset was split into 75% for training, 10% for validation and 15% for testing, the dataset was vectorized and fed into the model, which learns the patterns within the dataset. The dataset was obtained from Kaggle.com (Table 3.2). A learned model is dumped to the system's disk after training.

The model was trained using the essential variables with respect to the training from the datasets. The values of the X variable (train variable) are from: 'bvn', 'nin', 'typeofaction', 'sourceid', 'destinationid', 'amountofmoney' while the values of the Y variable (predict variable) used for

money laundering prediction is from ‘isfraud’. For money laundering detection, the Y variable (predict variable, also called the label) is ‘isFlagged’.

After the model training, the model was pickled to enable deployment on the development server using the flask (a python framework for website development).

iv. Advantages of the Proposed System

The advantages of the proposed system are listed below:

a) More accurate prediction due to balance dataset and removal of data without label.

b) The model was deployed on a localhost as a web app, hence to use the model requires no expertise of a data scientist or a machine learning engineer.

c) A user-friendly web interface.

3.4 Model Analysis

This model was built to run on the web, on a very low computing system and cross-platform on Linux, Windows, and Mac that meets the minimum system specification.

The model was tested on a Windows 10 computer with a minimum RAM of 1 gigabyte and a minimum storage requirement of 2 gigabytes. However, the model was built on a Windows 10 computer with a specification below.

Table 3.1 Test Device Specification

Requirement	*Minimum Specification*	*Built System Specification*
System OS	Windows/Linux /Mac	*Windows 10*
Platform	Localhost/Internet	Localhost
Browser	Any preferably Chrome, Opera, Avast Secure	Avast Secure and Chrome
Memory (RAM)	Minimum: 1 GB	8 GB
Free Storage Capacity	10 GB	35 GB

3.5 Analysis Tools

The research instrument (questionnaire) was analysed using simple count and percentage, and the model was analysed and evaluated using the classification metrics, such as accuracy, precision, recall, F-Measure and specificity, in sklearn python framework (library). Also, a confusion matrix was plotted for the model.

3.6 Model Development

The processes underwent for the model building will be discussed in subsections.

3.6.1 Model Architecture

A machine learning model for money laundering detection and prevention using financial data retrieved from Kaggle.com is presented in this study. The architecture of this model is shown in Figure 3.1. Initialising the development process, data has to be pulled from Kaggle.com (a storage repository for datasets amongst others). The format of the datasets was in comma separated value (csv) format. The datasets were imported using python's pandas's library. The various datasets were converted to vectors (dataframe), important features were extracted from the datasets then

concatenated into a dataset. The format for training the model is a vector format. The model was hyperparameter tuned till the model was ready for use. Finally, the model was saved to the computer disk, integrated into a web application and deployed to the web.

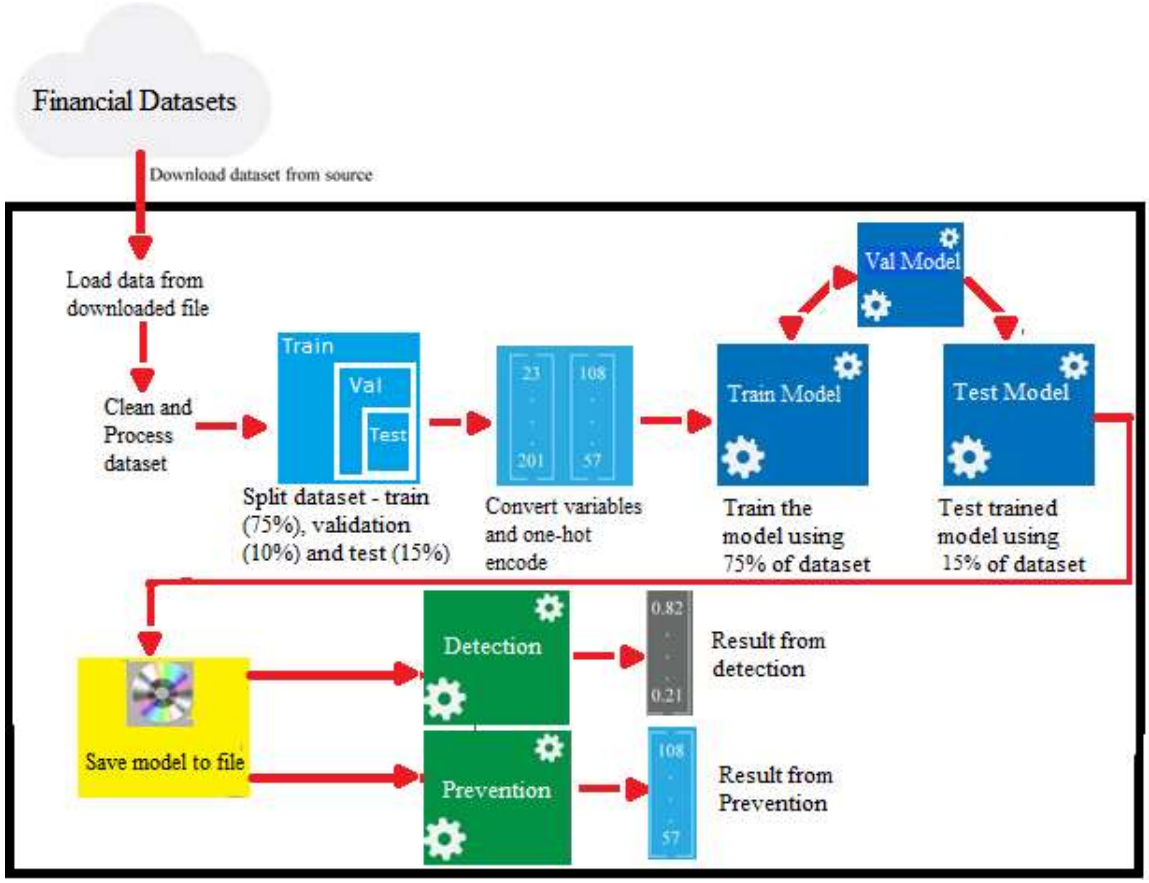


Figure 3.1 Model Architecture

Figure 3.1 depicts the stages undertaken to build the models. After the downloading, the dataset was cleaned, split into three sets namely train, validation and test. After training the model, it was saved as a pickled file which was used for the web application.

3.6.2 Data Acquisition

A well-presented and adequately balanced dataset is a herculean task; however, it is very necessary. A Nigerian financial dataset was not available online and no financial institution nor

financial regulatory body in Nigeria was willing to oblige our request for financial dataset. Hence, data were obtained from Kaggle.com, which is an open-source website. Financial data were collected via download from this site and were processed using pandas. Before the model was trained, two sets of random numbers were generated representing bank verification number (BVN) and national identity number (NIN). All the selected features (variables) from the datasets were concatenated to the BVN and NIN. This was done to include the Nigerian feature in the final dataset before training the model.

The datasets were downloaded from Kaggle.com with the details as indicated in Table 3.2.

Table 3.2 Source Details of Datasets

Filename	**URL**	**Date Downloaded**	**Records**	**Features**	**Label$_0$**	**Label$_1$**
PS_20174392719_1491204439457_log	https://www.kaggle.com/ealaxi/paysim1	*20-09-2019*	*28276056*	*11*	*1047433*	*1142*
MLtag	https://www.kaggle.com/maryam1212/money-laundering-data	16-04-2020	2340	8	941	1399
ML	https://www.kaggle.com/maryam1212/money-laundering-data	25-08-2021	73208391	3	448	1036

Label$_0$ stands for the number of records labelled as genuine while label$_1$ are the numbers of records labelled as money laundering.

3.6.3 Building the Model

Financial data were obtained from Kaggle.com during the development of the model and the fields not relevant to the prediction model of this research were stripped off. The stripped dataset was converted to vectors and trained using KNeighborsClassifier from the Sklearn library, which was then exported to deploy and use outside the development environment. The model hyperparameters were tuned for the model to achieve the best accuracy, and this continues iteratively till the training is completed, and the model is saved to disk, which can be deployed on several platforms for predicting and detection of money laundering. Figure 3.2 depict the process of training the model. The raw datasets were processed to a dataset and saved. The processed dataset was used to train the model. Note however it was 75% of the processed dataset that was used to train the model.

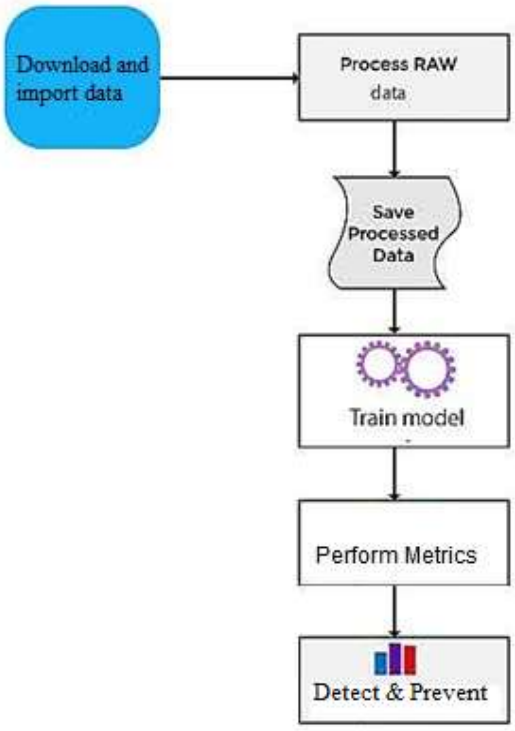


Figure 3.2 Training the detection model

3.6.4 Feature Engineering

The dataset (see Appendix A) used for the development of this model consist of both numeric and categorical data. The input (X variable) (see Appendix A) consists of a categorical data in the feature called ‘typeofaction’ with categories as ‘cash-in’ and ‘transfer’. A categorical data is one stored as a label instead of numeric; hence the categorical data was converted to numeric using one-hot encoding. Hence, ‘cash-in’, ‘transfer’ is represented as [1,0], [0,1] using one-hot encoding (see Appendix B).

3.6.5 Data Preprocessing

Data values as null and incorrect in were checked for during data processing and they were removed or replaced with the mean if the data value is numeric, such that there was no much loss in data. Pandas’ library function namely isna() and fillna() functions were used for the data processing (see Appendix B).

3.6.6 Modelling

Scikit learn was used for the modelling whereby the kNN algorithm was used (Algorithm 3.2). The model (Equation 3.1) has a fit function that accepts X as input and y as label (target). The model iterates over the processed data to learn patterns which are then saved as a model for detection and prediction.

3.6.7 Validation and Hyperparameter Tuning

After training the model (with 75% of the dataset), the validation data (10% of the dataset) was fed into the model to determine how it performs against this new data. A poor performance during testing after performing well during training means that the model suffers from overfitting. If the

model doesn't perform well even on the training data, then the model suffers from underfitting. Hence, we have to tune the hyperparameters like the learning rate, optimizer, or just get more data. This continues till we achieve good results from the model. But if it performs well during validation, it means the model is ready for testing where it is tested against a test set and real-world data to see how it performs.

3.6.8 Detection

When the model is done from being trained, validated, and tested with good results for accuracy, precision, and recall, it is then released for the use of detecting money laundering using the 'isfraud' feature of the dataset.

3.6.9 Prevention

A model for prevention was also developed using the 'isFlagged' features of the dataset in conjunction to the rules of CBN on account tiers. After training, validating and testing the prevention model, and having shown good results for the metrics, the preventive model was released.

3.7 Model Design

The model was designed using the kNN algorithm (page 63). The dataset was split into train and test using 75% and 25% respectively. The features of the train set and test set were scaled using the standard scaler function. After the model was scaled, the kNN algorithm was used with the following parameters; n_neighbors = 5, metric = 'minkowski', p = 2; for the algorithm to apply the Euclidean distance. The input's train set and the label's train set was fed into the specified kNN algorithm.

The trained model was made a pickle file (a python serialisation file) and dumped, so that it can be deployable. The test set was used to test the trained model and the metrics were obtained from the model's confusion matrix.

$$F: X \rightarrow y \; \forall \{X_i y_i\} \qquad i=1,2,3 \ldots n \tag{3.1}$$

Equation (3.1) is the model's mathematical equation where by F stands for the kNN model, X is for the input data, y is for the label. The model is the function that when given X, it gives y.

3.7.1 Input Design

The dataset's features were not all used as input data (X variable) to the model. The input data to the model were "typeofaction" and 'amountofmoney' for the detection model and 'type' and 'amount' for the prevention model.

Tabel 3.3 displays the input data which is also known as the X variable to the detection and prevention model.

Table 3.3 Input (X) Data

Feature	Description	DataType
	Detection Model	
typeofaction	The transaction type either deposit, transfer or withdrawal	String/categorical
amountofmoney	The amount of money in the transaction	Numerical
	Prevention Model	
Type	The transaction type either deposit, transfer or withdrawal	String/categorical
Amount	The amount of money in the transaction	Numerical

3.7.2 Interface Design

The interface design consists of web forms and input html tags as shown in Table 3.4. There are five (5) web pages which consist of HTML tags, styled with Bootstrap which is a CSS framework.

Table 3.4 Interface Design

Web Page	HTML Tags	Use
Index.html	Label, anchor and buttons	For description and hyperlink
Login.html	Label, text input, password input, button and anchor	For Username and Password input and submission
Home.html	Label, anchor and buttons	For description and hyperlink
Detection.html	Label, Text input, select input, radio input, button and anchor	For BVN. NIN, Account Number, Transaction Type, Amount, Previous Balance, IsFraud; submission and hyperlink
Prevention.html	Label, Text input, select input, radio input, button and anchor	For BVN. NIN, Account Number, Transaction Type, Amount, Previous Balance, IsFlagged; submission and hyperlink

3.7.3 Program Design

The program was designed using the python programming language. Data science and machine learning frameworks that were used in the development of the model are numpy, pandas, matplotlib, seaborn, pandas profiling and sklearn. Other python modules that were utilized are

random, pickle, os.path and datetime. In the sklearn framework, modules used are metrics, preprocessing, model_selection and neighors. The web-based was designed using HTML, CSS, Flask (a python framework for light web applications), Jinga. Numpy and Pandas were used for numerical and statistical computing. Matplotlib, seaborn and pandas profiling were used for visualization and descriptive statistics. Sklearn (Scikit-learn) was used for the model buiding. The model was picked using the pickle modules of python.

3.7.4 Process Design

The steps or procedures applied for the development of the models is given in Algorithm 3.1

Algorithm 3.1 Model's Procedure

Step 1. Begin

Step 2. Import python modules and data science modules

Step 3. Read dataset

Step 4. Perform data wrangling

Step 5. Clean and process dataset

Step 6. Perform exploratory and descriptive on dataset using pandas profiling

Step 7. Import machine learning modules

Step 8. Features extraction

Step 9. Features selection (Input and Label)

Step 10. Perform one hot encoding for categorical data

Step 11. Split dataset into train and test, that is, input train, input test, label train, label test

Step 12. Scale features

Step 13. Select parameter for kNN

Step 14. Hypertune parameters

Step 15. Fit input train and label train into model

Step 16. Pickle model

Step 17. Perform metrics (confusion matrix)

Step 18. Perform visualisation of train set

Step 19. Perform visualisation of test set

Step 20. Visualise confusion matrix

Step 21. Calculate metrics

Step 22. End

kNN is a supervised learning techniques for classification of data points of a given category with respect to the training set. The kNN algorithm applied for the model is given in

Algorithm 3.2 kNN Algorithm (English)

Step 1. Begin

Step 2. Load data

Step 3. Choose k

Step 4. For each observation (data point) in dataset do:

a. Calculate distance test data and each observation using Euclidean distance in **Error! Reference source not found.**

b. Sort the calculated distance in (a) in ascending order

c. Choose top k row from (b)

d. Assign class to test point based on number of k

Step 5. End

Algorithm 3.3 kNN Algorithm (Python)

```
Step 1.   import collections.Counter
Step 2.   import math
Step 3.   def(knn(data, query, k, distance_fn, choice_fn):
Step 4.       neighbor_distances_and_indices = []
Step 5.       for index, example in enumerate(data):
Step 6.           distance = distance_fn(example[:-1], query)
Step 7.           neighbor_distances_and_indices.append((distance, index))
Step 8.       sorted_neighbor_distances_and_indices = sorted(neighbor_distances_and_indices)
Step 9.       k_nearest_distances_and_indices = sorted_neighbor_distances_and_indices[:k]
Step 10.      k_nearest_labels = [data[i][-1] for distance, i in k_nearest_distances_and_indices]
Step 11.      return k_nearest_distances_and_indices , choice_fn(k_nearest_labels)

Step 12.  def mean(labels):
Step 13.          return sum(labels) / len(labels)

Step 14.  def mode(labels):
Step 15.      return Counter(labels).most_common(1)[0][0]

Step 16.  def euclidean_distance(point1, point2):
Step 17.      sum_squared_distance = 0
Step 18.      for i in range(len(point1)):
Step 19.          sum_squared_distance += math.pow(point1[i] - point2[i], 2)
```

Step 20. return math.sqrt(sum_squared_distance)

The formula to calculate the Euclidean distance is given in Equation (3.2)

$$d(x, y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \quad (3.2)$$

Equation 3.2 calculates the Euclidean distance where x and y are two points in Euclidean n-space, x_i and y_i are Euclidean vectors, starting from the origin of the space (initial point) and m is n-space

3.7.5 Database Design

The database used is SQLite (SQLAlchemy) of the flask server. It is a database toolkit for Python and also an Object Relational Mapper which writes SQL queries. The tables created are users and history. The user table stores the basic data for user that enables them to login to the system. The history table keeps history of each detection and prevention that was done by the system.

Table 3.5 Database Design

Table	Fieldname	Datatype	Size
user	name, username, password	string, string, string	15, 10, 50
history	nin, bvn, account_number, source_id, transaction_type, amount, previous_balance, isFraud, isFlagged	integer, integer, integer, integer, string, integer, integer, integer, integer	11, 11, 11, 20, 10, 20, 1,1

Table 3.5 illustrates the database tables. The field names, datatypes and data size of the tables was given in the respective columns.

3.7.6 Output/Report Design

After detection or prevention, the user will be presented with messages on the respective interface. The output forms are dynamic both in contents and in the colour scheme. If money laundering activity is detected or suspected, then red colour is presented in the background of the prediction report, else green colour becomes the background colour in the prediction report. The sample interfaces of the deployed web app were designed using HTML, CSS, Jinja and bootstrap (see Appendix C).

CHAPTER FOUR

RESULT AND DISCUSSION

4.1 Model Detection and Prevention

In following subsections, the two anti-money laundering models (for detection and prevention) built will be discussed.

4.1.1 Money Laundering Detection

The feature used for the training the model of money laundering detection consists of 'bvn', 'nin', 'typeofaction', 'sourceid', 'destinationid', 'amountofmoney', 'date', 'isfraud', 'typeoffraud', 'guiltyid', 'levelofcrime'. The label is ‘isfraud’ and this was used to train the model on detecting a transaction that is suspected as money laundering. Others features excluding the label makes up the input data.

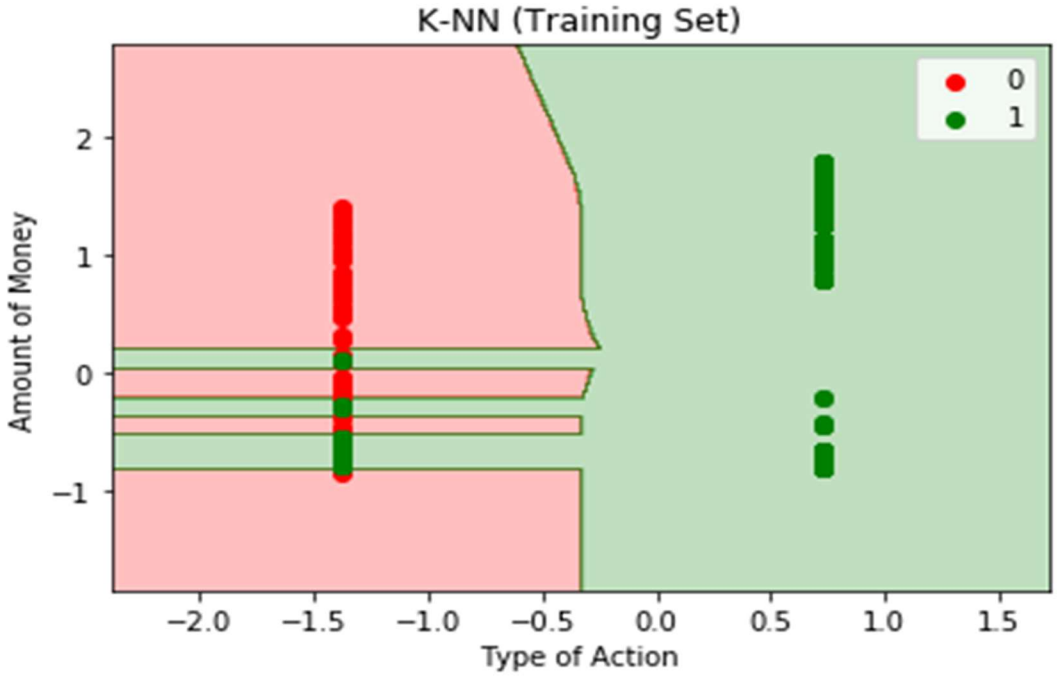


Figure 4.1 Visualization of the Detection Training Set

Figure 4.1 showed two categories based on the label. The ‘0’ stands for a transaction that is not money laundering while ‘1’ stands for transaction that are grouped as money laundering.

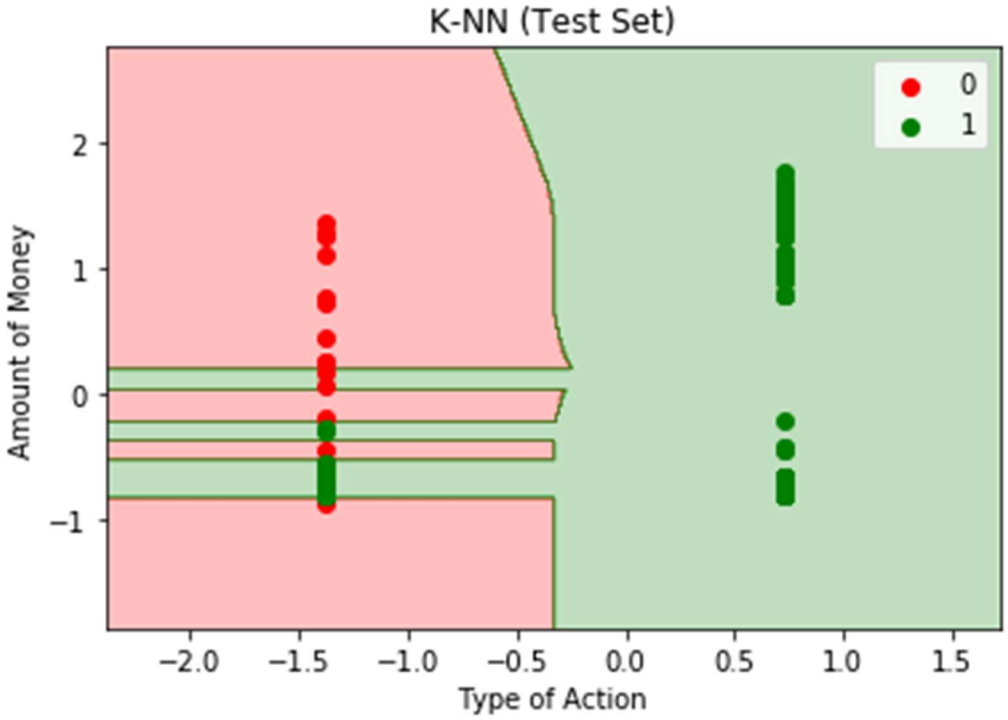


Figure 4.2 Visualisation of the Detection Test Set

After the model was trained, it was tested with 15% of the dataset. The visualisation on the test dataset is shown in Figure 4.2. It can be judged from the figure that the test model did well enough with an accuracy of 0.984 (that is 98.4%).

4.1.2 Money Laundering Prevention

The model for money laundering prevention was trained with features such as 'bvn', 'nin', 'type', 'nameOrig', 'nameDest', 'amount', 'date', "isFraud", "isflagged". The label was "isflagged" and the input data consist of other features excluding the label. The training was done with 75% of the dataset.

In addition to the trained prevention model, the Central Bank of Nigeria rules on the tiers of account was automated programmatically using nested if and if else statements. The spinet of the code is presented in Program 1.

```
def rule(tier, amount, balance):
    account_type = tier
    amount = amount
    balance = balance
    account_rules = {1:[20000, 200000], 2:[50000, 400000], 3:[None, None]}
    status = False
    if account_type == 1:
         if amount > account_rules[account_type][0] or balance >
            account_rules[account_type][1] or (amount + balance) >
            account_rules[account_type][1]:
           status = True
    elif account_type == 2:
        if (amount > account_rules[account_type][0] or balance >
            account_rules[account_type][1] or (amount + balance) >
            account_rules[account_type][1]):
           status = True
    elif account_type == 3:
        status = False
    else:
        status = None
    if status == True:
        return 1
    elif status == False:
```

```
        return 0
    else:
        return 2
if status == 1:
      msg = f'hello, your class is {predClass} with probability of {pred_given}
      based on the imputed features. Your transaction won\'t proceed!'
      return render_template('/prevented.html', data1=features, result=output,
      message=msg, colour=colour, bg = bg)
elif status == 0:
      msg = f'Hello, your class is {predClass} with probability of {pred_given}
      based on the imputed features. Your transaction will proceed!'
      return render_template('/prevented.html', data1=features, result=output,
      message=msg, colour=colour, bg = bg)
else:
      msg = f'Hello, your class is {predClass} with probability of {pred_given}
      based on the imputed features. I don\'t understand your input. Try again!'
      return render_template('/prevented.html', data1=features, result=output,
      message=msg, colour=colour, bg = bg)
```

Program 1 Spinet of Prevention Code

4.2 Results of Model Training and Testing

The metrics from training and testing the detection and prevention models are presented in the following subsections.

4.2.1 Money Laundering Detection

The detection model did well with the following: observations detected as True Negative (TN) is 13, observations detected False Positive (FP) is 6, observations detected as False Negative (FN) is 0 and observations detected as True Positive (TP) is 351. The confusion matrix of this model is given in Figure 4.3:

```
[[ 13   6]
 [  0 351]]
```

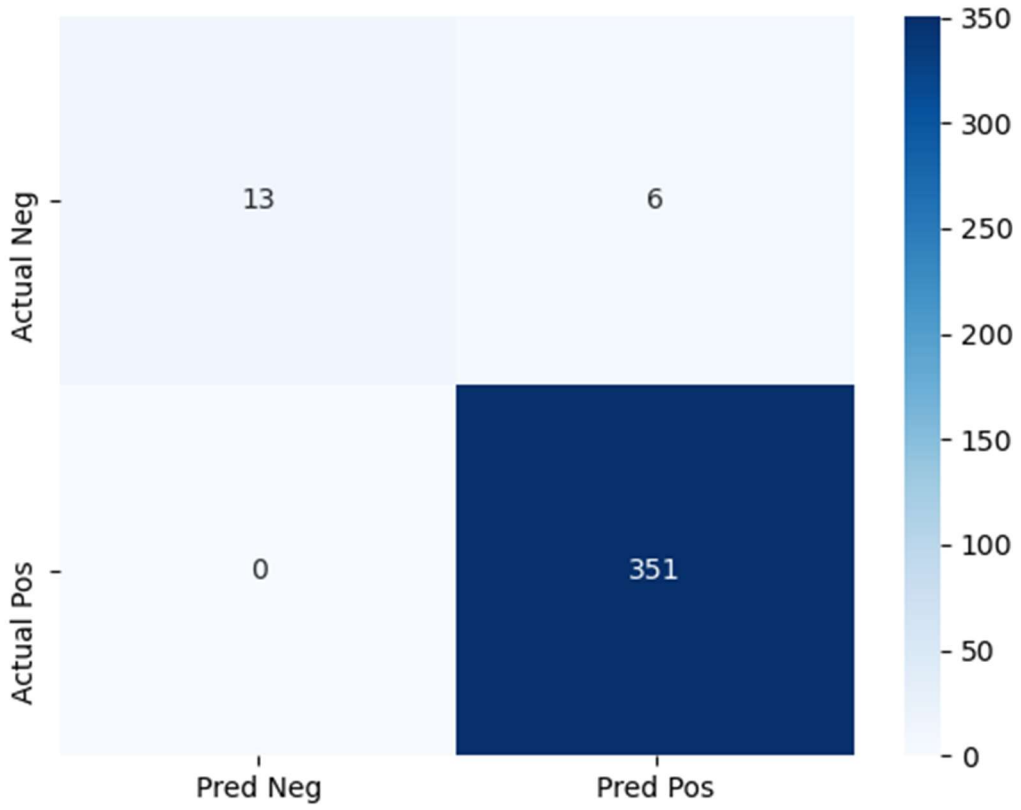


Figure 4.3 Confusion Matrix of Detection Model

The following metrics were obtained from the model:

Table 4.1 Metrics for Detection Model

Accuracy	Precision	Recall	F-Measure	Specificity
0.984	0.983	1.000	0.992	0.992

Table 4.1 depicts an accuracy of 98.4%, precision of 98.3%, recall of 100%, F-Measure and specificity of 99.2%. These metrics were calculated in jupyter notebook (see Appendix D).

4.2.2 Money Laundering Prevention

The prevention model, from the test set detected 4 transactions as True Negative (TN), 58 transactions as False Positive (FP), 9 transactions as False Negative (FN) and 299 transactions as True Positive (TP). The confusion matrix of this model is in Figure 4.4:

```
[[  4  58]
 [  9 299]]
```

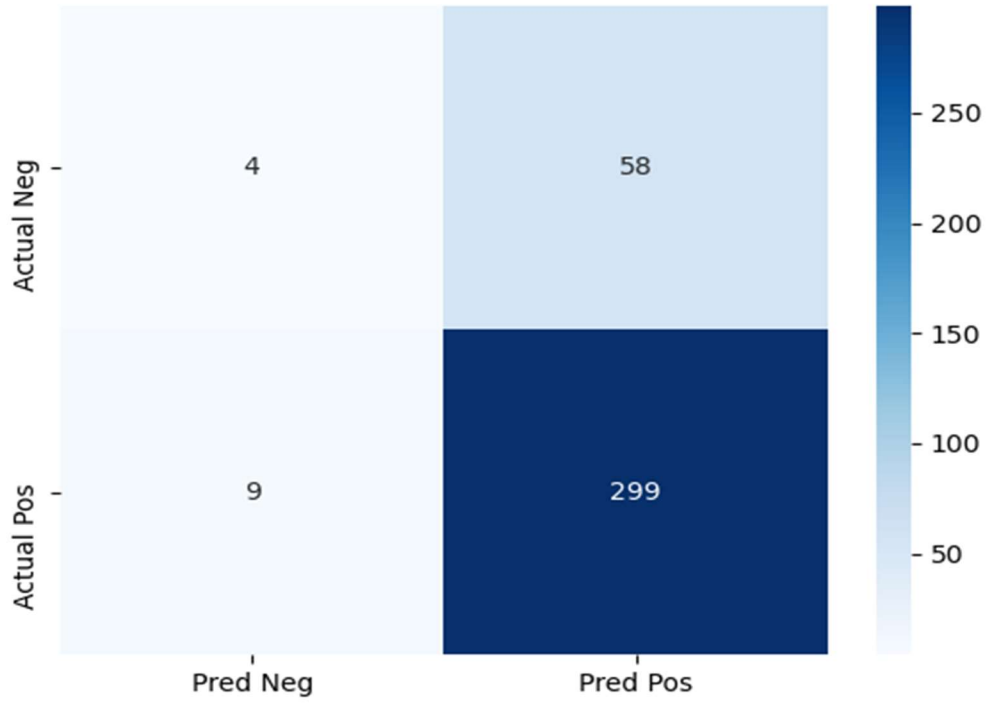


Figure 4.4 Confusion Matrix of Prevention Model

The following metrics were obtained from the model:

Table 4.2 Metrics for Prevention Model

Accuracy	Precision	Recall	F-Measure	Specificity
0.819	0.838	0.971	0.899	0.899

Table 4.2 depicts an accuracy of 81.9%, precision of 83.8%, recall of 97.1%, F-Measure and specificity of 89.9%

4.3 Model Evaluation

The performance of the new model was compared with the existing model as shown in Table 4.3. The visualisation of the models' performance comparison is presented in Figure 4.5

Table 4.3 Performance Evaluation

	F1 score	Precision	Recall	Accuracy
Existing Model	0.97	0.97	0.98	0.92
Our Model (D)	0.99	0.98	1.00	0.98
Our Model (P)	0.90	0.84	0.97	0.82

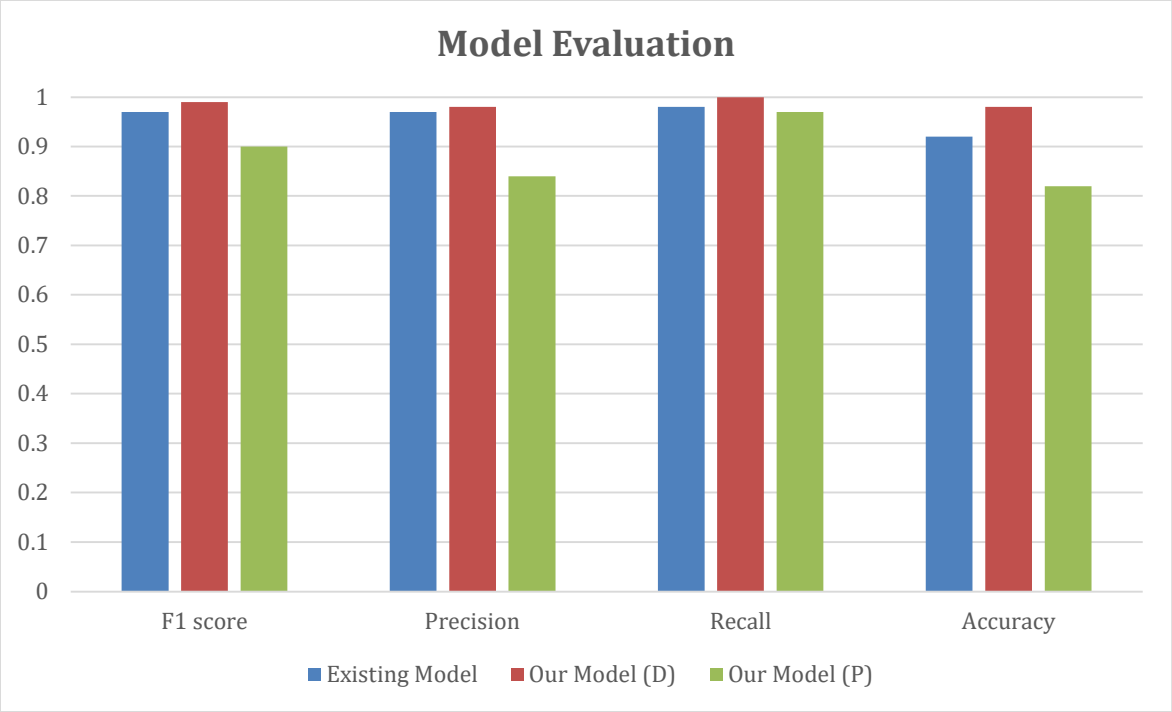


Figure 4.5 Model Evaluation

Model (P) denotes the prevention model while Model (D) denotes the detection model. The metrics of the detection model outperform that of the existing system.

4.4 Statistical Analysis and Finding

The questionnaire received poor response from the respondents, possible reasons might be the confidentiality and secrecy associated to financial matters, also the lack of institution's email address for the researcher. Hence, the questionnaire sent via google form was not filled and return by the recipients.

Table 4.4 Gender Analysis

Gender	Count	Percentage
Male	4	80%
Female	1	20%

Table 4.4 shows that only a female (representing 20%) was among the respondents, while four male which represents 80% responded to the questionnaire. The questionnaire is not gender based and it was administered randomly.

Table 4.5 Profession (Role) Analysis

Profession (Role)	Count	Percentage
Customer Service Branch Manager	2	40%
Operations Personnel	2	40%
Branch Manager	1	20%

Table 4.5 shows that a branch manager, two operations personnel and two customer service branch manager which represents 20%, 40% and 40% respectively answered the questionnaire. The questionnaire was administered randomly but only to those whose role in the bank is concerned to the area of research.

Table 4.6 Experience Analysis

Experience	Count	Percentage
5 – 10 years	2	40%
11 – 15 years	0	0
16 – 20 years	3	60%

The experience of the respondents as showed in Table 4.6, falls under two class; two respondents (40%) have experience between 5-10 years, while three respondents (60%) have experience between 16-20 years.

Respondents' response to Section B of the research instrument (questionnaire) communicates the following:

1. Money laundering detection in Nigeria is automated, however, the system needs improvement because it is rule based and not self-learning.
2. The channels used for money transfer in Nigeria are: USSD, NIP, NAPS, online banking, bank's app.
3. There is no uniform (single) gateway for all money transfer in Nigerian irrespective of the channel of transfer.
4. The different account types and their limits as follows:

Table 4.7 Account Types and Limits

Account Type	Limits		
	Single Deposit	Daily Transaction	Maximum Balance
Tier 1	20,000	30,000	200,000
Tier 2	50,000	100,000	400,000
Tier 3	None	None	None

5. Transaction limits are subject to account type (Table 4.7) and not the Bank Verification Number (BVN).
6. Money transferred from outside Nigeria to a Nigerian account is subject to the banking laws in Nigeria.
7. The following are possible reason an account can be flagged suspicious to money laundering:
    a. Amount of cashflow
    b. Sudden change in transaction history/pattern
    c. Transfers from various senders over a short duration
    d. Recipient of fund being unable to state a legitimate relationship between him and the sender.
    e. Exceeding the maximum limit flag.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This research developed a model that aids in making better financial transactions decisions through the detection and prevention of money laundering. The model is a statistical model trained on the data from Kaggle.com, where the model learns predictive features related to financial change. Statistical survey was done before the building of the model, so as to analyse the present system and obtain data from experts in the field.

The model was built using Python and a couple of machine learning libraries. Sklearn (scikit-learn) was used to create the machine learning model, NumPy was used for vector processing, pandas to handle Comma Separated Value (CSV), the format of the dataset.

From the questionnaire administered during this study, it is clear that there is a need for a better model and system for money laundering detection. As seen in the previous chapters, a majority of the surveyed population believe theirs is a need for a smarter money laundering detection model.

The model on prevention was able to prevent money laundering using the financial transaction limits as given by the Central Bank of Nigeria in conjunction to the 'isFlagged' feature of the dataset.

5.2 Recommendations

The following are recommendations made due to the discoveries during this study.

1. The government should embrace smarter technology such as this, to boost money laundering detection and prevention
2. The government should revise their anti-money detection laws and the bank verification number (BVN) law, so that the money laundering check will be on BVN and not on account numbers.
3. Grants can be given by the government to support researchers in this domain so that the resources necessary to gather data and build models will be less of a burden on the developer.
4. The government could be or rent out computing space for researchers that are carrying out research that is beyond the capacity of a personal computer.

5.3 Suggestions for Further Studies

In order to expand human knowledge, there is a need to highlight some suggestions based on this research. The recommendations for further studies are as follows;

1. Further research can be carried out using TensorFlow or a deep neural network.
2. Deployment of the model on a mobile platform should be considered by future researchers.
3. Data sources from less developed countries can be used also and be added to improve the performance of the model in such regions.
4. Further research could seek to deploy the model on edge devices like Arduino and Raspberry Pi, where the sensors can be used to read inputs directly from the fields, which are cheap and affordable.

5.4 Contributions to Knowledge

The following areas are the contribution to knowledge made by this study

1. The study ascertained that kNN which is a supervised machine learning algorithm had an improved recall of 100% against the 97% of the existing system (Figure 17 and Table 4.3).

2. This study designed and deployed two machine learning models for money laundering detection and money laundering prevention.

3. The study ascertained that the built detection model performs better than the existing system as shown in the test metrics – recall, f1-score/f1-measure and precision.

REFERENCES

Ahmed, T. I. (2019). Money Laundering and Financial Crimes in Nigeria. *IOSR Journal of Economics and Finance*, 47 (4), 61-69

Alarab, I., Prakoonwit, S., and Nacer, M. I. (2020). Comparative Analysis Using Supervised Learning Methods for Anti-Money Laundering in Bitcoin, 3, 11-17.

Alexandre, C., and Balsa, J. (2016). Client Profiling for an Anti-Money Laundering System.

Alotibi, J., Almutanni, B., Alsubait, T., Alhakami, H., and Baz, A. (2022). Money Laundering Detection using Machine Learning and Deep Learning. In *IJACSA) International Journal of Advanced Computer Science and Applications*. 13(10), 732-738. www.ijacsa.thesai.org

Banerjee, R., Bourla, G., Chen, S., and Purohit, S. (2018). Comparative Analysis of Machine Learning Algorithms through Credit Card Fraud Detection. 2018 IEEE Symposium on Computational Intelligence. 12(11), 18--21.

Braun, S., Glück, T., and Röglinger, M. (2018). Towards a digitalization framework for regulatory technology. Journal of Information Technology, 33(4), 283-298.

Broussard, R. P., and Wey, W. M. (2020). Machine Learning for Anti-Money Laundering: Can Supervised Learning Be Trusted? Expert Systems with Applications, 139, 112847.

Brust, K. (2019). Anti-Money Laundering in 2019: 6 Emerging Compliance Trends. Journal of Money Laundering Control, 22(2), 178-183.

Carlos, C., and Steven, M. (2017). *Database Systems: Design, Implementation, and Management* (12th ed.). Cengage Learning.

CBN (2019). CBN amends anti-money laundering laws - Punch Newspapers.

CBN (2021a). Administrative Sanction Regime.

CBN (2021b). CBN AMENDS ITS LAWS ON ANTI-MONEY LAUNDERING.

CBN (2021c). CBN Financial Inclusion Drive - Tiered know Your Customer Strategy.

CBN (2021d). CBN goes tough on money laundering with new rules.

CBN (2021e). QUICK REVIEW OF THE CBN'S ANTI-MONEY LAUNDERING_COMBATING THE FINANCING OF TERRORISM (AML_CFT) POLICY AND PROCEDURE MANUAL.

Cem, D. (2023). Anti Money Laundering Algorithms in 2023: Tackling AML with AI. *AI Multiple*, 1--6.

Colombo, G., and Aicardi, G. (2021). The Role of FinTech in Preventing Money Laundering: A Scoping Review. Technological Forecasting and Social Change, 171, 120939.

Dalpiaz, F. (2020). *Research Challenges in Information Science*. https://doi.org/10.1007/978-3-030-50316-1

Demetis, D. S. (2018). Fighting money laundering with technology: A case study of Bank X in the UK. *Decision Support Systems*.

Doppalapudi, P. K., Pankaj, K., Adrian, M., Christophe, R., Rick, S., Scott, W., and Shuo, Z. (2022). The fight against money laundering: Machine learning is a game changer. *McKinsey Global Publishing, McKinsey and Company*, *October*.

Doppalapudi, P. K., Pankaj, K., Adrian, M., Christophe, R., Rick, S., Scott, W., and Shuo, Z. (2022). The fight against money laundering: Machine learning is a game changer(October).

EFCC (2019). HIGH PROFILE CASES BEING PROSECUTED BY THE EFCC FOR AG.

EFCC (2021). MONEY LAUNDERING CASES IN NIGERIA.

Elham, N. S., Mustapha, A., and Hassan, M. S. (2019). A review of network-based money laundering detection techniques. Journal of Money Laundering Control, 22(4), 512-530.

Enofe, A. O., Aliu, A. K., and Ombu, A. (2018). Money Laundering and The Nigerian Economy. *International Journal of Advanced Academic Research*.

Frumerie, R. (2021). Money Laundering Detection using Tree Boosting and Graph Learning Algorithms.

Ghosh, A., and Chaudhary, N. (2020). Digital Identity and Money Laundering: An Empirical Analysis. Journal of Financial Crime, 27(3), 951-968.

I-Hsien Ting, Hui-Ju Wu, and Tien-Hwa Ho (2010). Mining and Analyzing Social Networks (Studies in Computational Intelligence, 288). *Springer-Verlag Berlin Heidelberg*.

Joana, F. O. M. (2015). Risk Analysis in Money Laundering: A Case Study.

Jullum, M., Løland, A., and Huseby, R. B. (2020). Detecting money laundering transactions with machine learning, *23*(1), 173--186. https://doi.org/10.1108/JMLC-07-2019-0055

Kabir, M. N., and AlJamea, M. M. (2021). The Role of Biometrics in Anti-Money Laundering: A Review and Research Agenda. Journal of Money Laundering Control, 24(4), 652-673.

Kamps, J., and Klein, T. (2020). Money Laundering in Cryptocurrencies: An Emerging Threat. Journal of Financial Crime, 27(3), 938-950.

Kateryna, C. (2017). *Machine Learning Methods for Malware Detection and Classification.*

Khan, S., and Verma, N. (2021). Big Data Analytics in Anti-Money Laundering: A Comprehensive Review. Expert Systems with Applications, 182, 115842.

Kharote, M., and Kshirsagar, V. P. (2014). Data Mining Model for Money Laundering Detection in Financial Domain. *International Journal of Computer Applications*, *85*(16).

Kingston, K. G. (2020). Concealment of illegally obtained assets in {Nigeria}: {Revisiting} the role of the churches in money laundering. *African Journal of International and Comparative Law*, *28*(1), 106--121. https://doi.org/10.3366/ajicl.2020.0304

Kuda (2021). Account Levels, Rules and Limits _ Kuda Help Center.

Le Nhien, A. K., Sammer. Markos, and M-Tahar, K. (2010). A data mining-based solution for detecting suspicious money laundering cases in an investment bank.

Li, J., and Shi, Y. (2021). A Survey on Machine Learning for Anti-Money Laundering. IEEE Transactions on Computational Social Systems, 8(3), 683-697.

Liu, B., Zheng, M., and Choo, K. K. R. (2020). A survey of blockchain technology for anti-money laundering. IEEE Transactions on Computers, 69(3), 413-427.

Llu´ıs, A., Awasthi, A., and Jorg, L. (2012). Genetic Clustering Algorithms for Detecting Money-Laundering.

Lokanan, M. (2022). Predicting Money Laundering Using Machine Learning and Artificial Neural Networks Algorithms in Banks. *Applied Security Research*, *August*. https://doi.org/10.1080/19361610.2022.2114744

Lokanan, M. (2022). Predicting Money Laundering using Machine Learning and Artificial Neural Networks Algorithms in Banks. *Applied Security Research.* Advance online publication. https://doi.org/10.1080/19361610.2022.2114744

Lopez-Rojas, E. A., and Axelsson, S. (2012). Money Laundering Detection using Synthetic Data.

Manjunath, K. V. (2015). Data Mining Techniques for Anti Money Laundering. *International Journal of Advanced Research in Science,*

Mayhew, H., Bond, C., and Rai, J. (2021). The impact of evolving AML regulations on money laundering detection and prevention. International Journal of Law, Crime, and Justice, 63, 100612.

NBS (2016). CRIME STATISTICS: Nigerian Prisons.

NBS (2017). Crime Statistics: Reported Offences by Type and State.

Nikolov, R., Sabol, T., and Munzert, S. (2022). The ethical dimensions of money laundering detection: A multi-disciplinary review. Computer Law and Security Review, 42, 105501.

Odii, *, Hampo, J. N., and Onwuama, N. F. O. (2019). COMPARATIVE ANALYSIS OF MALWARE DETECTION TECHNIQUES USING SIGNATURE, BEHAVIOUR AND HEURISTICS. *International Journal of Computer Science and Information Security (IJCSIS)*, *17*(7), 33. https://sites.google.com/site/ijcsis/

OECD. (2013). *Illicit Financial Flows from Developing Countries: Measuring OECD Responses.*

OECD. (2019). *Money Laundering and Terrorist Financing Awareness Handbook for Tax Examiners and Tax Auditors.*

Ogbodo, U. K., and Mieseigha, E. G. (2013). The Economic Implications of Money Laundering in Nigeria.

Rafał Drezewski, Grzegorz Dziuban, Łukasz Hernik, and Michał Paczek (2015). Comparison of Data Mining Techniques for Money Laundering Detection System.

Rafay, A. (2021). *Money Laundering and Terrorism Financing in Global Financial Systems* (Issue March). https://doi.org/10.4018/978-1-7998-8758-4

Rafay, A. (2021). *Money Laundering and Terrorism Financing in Global Financial Systems*. https://www.google.co.in/books/edition/Money_Laundering_and_Terrorism_Financing/XMIkEAAAQBAJ?hl=en&gbpv=0 https://doi.org/10.4018/978-1-7998-8758-4

Ramya, K., Prof, A., Sathak, M., and College, E. (2022). Comparative Analysis and Implementation of AI Algorithms for Money Laundering Detection. *Emerging Technologeis and Innovative Research*, *9*(8), 700–704.

Ramya, K., Shenija, S., Banu, S. S., Ramachandran, M. C., and Balamurugan, G. (2022). Comparative Analysis and Implementation of AI Algorithms for Money Laundering Detection. *Emerging Technologeis and Innovative Research*, *9*(8), 700--704.

Ruiz, E. P., and Angelis, J. (2022). Combating money laundering with machine learning – applicability of supervised-learning algorithms at cryptocurrency exchanges. *Journal of Money Laundering*, *25*(4), 766--778. https://doi.org/10.1108/JMLC-09-2021-0106

Sahadev, S., and Gupta, S. (2021). Anti-money laundering compliance and the role of technology: A systematic literature review and research agenda. Journal of Financial Crime, 28(1), 86-107.

Salehi, A., Mehdi, G., and Mohammed, F. (2017). Data Mining Techniques for Anti Money Laundering. *International Journal of Applied Engineering Research*, *12*(20).

Shaw, J., Bandyopadhyay, S., and Singh, V. (2023). Strengthening international collaboration in money laundering detection: Challenges and opportunities. International Journal of Criminology and Sociology, 2(1), 112-130.

Shcherbakov, M., and Smith, A. (2018). An Overview of Blockchain and Cryptocurrency: A Deeper Dive into Money Laundering Threats. Journal of Money Laundering Control, 21(1), 23-33.

Shijia, G., and Dongming, X. (2009). Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering. *Expert Systems with Applications*.

Shun, K., and Ryusuke, C. (2019). Detecting problematic transactions in a c2c ecommerce network.

Soltani, R., Nguyen, U. T., Yang, Y., Faghani, M., Yagoub, A., and An, A. (2016). A New Algorithm for Money Laundering Detection Based on Structural Similarity. Advance online publication. https://doi.org/10.1109/UEMCON.2016.7777919

Stojanović, J., and Milovanović, V. (2020). The Impact of FinTech on Money Laundering and Regulatory Practices. Economic Research-Ekonomska Istraživanja, 33(1), 3149-3163.

Sujith, A.L.N.V., Qureshi, N. I., Harshavardhan, V., Dornadula, R., Rath, A., Prakash, K. B., and Singh, S. K. (2022). A Comparative Analysis of Business Machine Learning in Making Effective Financial Decisions Using Structural Equation Model (SEM). *Hindawi Journal of Quality Food*.

Timm, F., Andrea Zasada, and Felix Thiede (2016). Building a Reference Model for Anti-Money Laundering in the Financial Sector.

UNODC (2017). UNITED NATIONS OFFICE ON DRUGS AND CRIME.

UNODC. (2011). *UNITED NATIONS OFFICE ON DRUGS AND CRIME Vienna Independent In-depth evaluation of The Global Programme against Money Laundering, Proceeds of Crime and the Financing of Terrorism*. www.unodc.org

Vitas, D., and Džemidžić, J. (2019). Money Laundering Detection through Data Mining Techniques. Journal of Money Laundering Control, 22(3), 425-437.

Xingqi, W., and Guang, D. (2009). Research on Money Laundering Detection Based on Improved Minimum Spanning Tree Clustering and Its Application. *Second International Symposium on Knowledge Acquisition and Modeling*.

Zhang, Y., Zhao, J., and Zhang, X. (2021). A survey of machine learning techniques for money laundering detection. Journal of Financial Crime, 28(1), 170-186.

Zhiyuan, C., van Dinh, K., Amril, N., Ee, N. T., and Ettikan, K. (2014). Exploration of the Effectiveness of Expectation Maximization Algorithm for Suspicious Transaction Detection in Anti-Money Laundering. *IEEE Conference on Open Systems (ICOS)*.

Appendix A

Sample Money Laundering Dataset

A1 - Detection Dataset

bvn	nin	typeofaction	sourceid	destinationid	amountofmoney	date	isfraud	typeoffraud	guiltyid	levelofcrime
22119875524	15530544386	cash-in	30105	28942	494528	19/07/2019 14:40	1	type1	30105	head
85297671377	12747842551	cash-in	30105	8692	494528	17/05/2019 14:57	1	type1	80740	head
21234559911	16692603144	cash-in	30105	60094	494528	20/07/2019 13:20	1	type1	92735	head
52114172303	43977997907	cash-in	30105	20575	494528	03/07/2019 14:15	1	type1	1615	head
19702361405	81375635965	cash-in	30105	45938	494528	26/05/2019 10:40	1	type1	4161	head
14523655048	47242239515	cash-in	30105	54971	494528	06/04/2019 11:18	1	type1	33203	head
14430431638	20699805190	cash-in	30105	62257	494528	23/04/2019 08:20	1	type1	41969	head
12147291447	26915114324	cash-in	30105	1020	494528	04/07/2019 08:59	1	type1	37177	head
66398911781	30149823720	cash-in	30105	98751	494528	06/05/2019 12:49	1	type1	31439	head
18835081725	15222135314	cash-in	30105	82016	494528	09/06/2019 12:27	1	type1	84352	head
67563051652	19559142182	cash-in	30105	13800	494528	17/04/2019 08:18	1	type1	95615	head
64629524002	10998243250	cash-in	30105	52681	494529	18/07/2019 13:57	1	type1	41220	head

13859521101	48093825349	cash-in	30105	80113	494529	27/03/2019 13:13	1	type1	39284	head
20368081116	35808742499	cash-in	30105	64316	494529	09/06/2019 11:43	1	type1	82688	head
14657772029	41732894302	cash-in	30105	94472	494529	21/02/2019 12:34	1	type1	30965	head
10429262002	40670662496	cash-in	30105	23762	494529	07/04/2019 12:43	1	type1	72032	head
14495031473	14973775399	cash-in	30105	12860	494529	28/04/2019 15:08	1	type1	24041	head
17781401100	13158075134	cash-in	30105	61008	494529	19/04/2019 15:02	1	type1	82410	head
45813381331	27665671830	cash-in	30105	3164	494529	19/05/2019 09:26	1	type1	24441	head
82093831491	71450867275	cash-in	30105	24259	494529	20/04/2019 11:48	1	type1	43439	head
80148841112	45192651484	cash-in	80740	29758	388294	03/05/2019 11:40	1	type1	69208	head
15488582054	59719803772	cash-in	80740	47869	388294	14/06/2019 08:10	1	type1	71475	head
68888131480	95238393360	cash-in	80740	79227	388294	05/03/2019 13:22	1	type1	75076	head
22177052226	20255194107	cash-in	80740	68225	388294	03/03/2019 09:24	1	type1	82930	head
14402881826	45591602642	cash-in	80740	65031	388294	10/03/2019 12:45	1	type1	13919	head
43173031245	23046001038	cash-in	35086	36839	426669 3	02/03/2019 09:32	0	none	88290	colleague

10368071831	25745301301	cash-in	45526	18743	395032	25/05/2019 13:49	0	none	77370	colleague
62914631546	22389855328	cash-in	99751	17283	613998	24/06/2019 09:45	0	none	58669	colleague
15839754335	30084313516	cash-in	5499	88631	315290	01/05/2019 12:40	0	none	728	colleague
38318811565	27322734159	cash-in	86622	61158	574899	16/03/2019 12:01	0	none	64353	colleague
10566731103	53098405411	cash-in	94485	13021	640350	23/05/2019 11:56	0	none	17894	colleague
15936311385	18450468625	cash-in	63307	35542	636031	29/03/2019 09:36	0	none	40237	colleague
78729891188	16283382415	cash-in	95376	59438	507286	19/06/2019 09:00	0	none	58854	colleague
17814951391	23451664519	cash-in	86807	13105	458217	03/04/2019 14:41	0	none	56874	colleague
22330571426	22967342697	cash-in	88920	71196	205967	07/05/2019 15:46	0	none	52734	colleague
17700612944	43417334645	cash-in	334	28048	462833	03/07/2019 12:28	0	none	32713	colleague
17439251631	42645634740	cash-in	17447	12872	179772	04/03/2019 09:12	0	none	71486	colleague
15999821385	26445834404	cash-in	98282	7583	121824	11/05/2019 10:38	0	none	20358	colleague
22433797798	47261951190	cash-in	31654	50710	336858	19/03/2019 08:50	0	none	77549	colleague
86228192256	30636247472	cash-in	11850	55033	404047	12/03/2019 09:47	0	none	92740	colleague

71995911730	52243853568	cash-in	32093	32298	2529811	06/06/2019 11:29	0	none	48631	colleague
20215821535	32119412331	cash-in	58094	16511	737413	27/02/2019 15:48	0	none	58492	colleague

A2 - Prevention Dataset

bvn	**nin**	**type**	**nameOrig**	**nameDest**	**amount**	**date**	**isFraud**	**isflagged**
22119875524	15530544386	PAYMENT	C1231006815	M1979787155	9839.64	19/07/2019 14:40	0	1
85297671377	12747842551	PAYMENT	C1666544295	M2044282225	1864.28	17/05/2019 14:57	0	1
21234559911	16692603144	TRANSFER	C1305486145	C553264065	181	20/07/2019 13:20	1	1
52114172303	43977997907	CASH_OUT	C840083671	C38997010	181	03/07/2019 14:15	1	1
19702361405	81375635965	PAYMENT	C2048537720	M1230701703	11668.14	26/05/2019 10:40	0	1
14523655048	47242239515	PAYMENT	C90045638	M573487274	7817.71	06/04/2019 11:18	0	1
14430431638	20699805190	PAYMENT	C154988899	M408069119	7107.77	23/04/2019 08:20	0	1
12147291447	26915114324	PAYMENT	C1912850431	M633326333	7861.64	04/07/2019 08:59	0	1
66398911781	30149823720	PAYMENT	C1265012928	M1176932104	4024.36	06/05/2019 12:49	0	1
18835081725	15222135314	DEBIT	C712410124	C195600860	5337.77	09/06/2019 12:27	0	1
67563051652	19559142182	DEBIT	C1900366749	C997608398	9644.94	17/04/2019 08:18	0	1
64629524002	10998243250	PAYMENT	C249177573	M2096539129	3099.97	18/07/2019 13:57	0	1
13859521101	48093825349	PAYMENT	C1648232591	M972865270	2560.74	27/03/2019 13:13	0	1
20368081116	35808742499	PAYMENT	C1716932897	M801569151	11633.76	09/06/2019 11:43	0	1
14657772029	41732894302	PAYMENT	C1026483832	M1635378213	4098.78	21/02/2019 12:34	0	1
10429262002	40670662496	CASH_OUT	C905080434	C476402209	229133.9	07/04/2019 12:43	0	1

14495031473	14973775399	PAYMENT	C7617507069	M1731217984	1563.82	28/04/2019 15:08	0	1
17781401100	13158075134	PAYMENT	C1237762639	M1877062907	1157.86	19/04/2019 15:02	0	1
45813381331	27665671830	PAYMENT	C2033524545	M473053293	671.64	19/05/2019 09:26	0	1
82093831491	71450867275	TRANSFER	C1670993182	C1100439041	215310.3	20/04/2019 11:48	0	1
80148841112	45192651484	PAYMENT	C20804602	M1344519051	1373.43	03/05/2019 11:40	0	0
15488582054	59719803772	DEBIT	C1566511282	C1973538135	9302.79	14/06/2019 08:10	0	0
68888131480	95238393360	DEBIT	C1959239586	C515132998	1065.41	05/03/2019 13:22	0	0
22177052226	20255194107	PAYMENT	C504336483	M1404932042	3876.41	03/03/2019 09:24	0	0
14402881826	45591602642	TRANSFER	C1984094095	C932583850	311685.9	10/03/2019 12:45	0	0
56832182976	93514985186	PAYMENT	C1043358826	M1558079303	6061.13	17/07/2019 15:53	0	0
23047941965	13003463988	PAYMENT	C1671590089	M58488213	9478.39	04/03/2019 08:11	0	0
17968634158	16884122206	PAYMENT	C1053967012	M295304806	8009.09	25/05/2019 14:38	0	0
10227462088	54636524223	PAYMENT	C1632497828	M33419717	8901.99	22/05/2019 12:52	0	0
81750881629	44295295432	PAYMENT	C764826684	M1940055334	9920.52	06/05/2019 15:18	0	0
16326791314	86094941269	PAYMENT	C2103763750	M335107734	3448.92	26/06/2019 13:33	0	0
19376737348	16387102265	PAYMENT	C215078753	M1757317128	4206.84	27/04/2019 10:52	0	0
21670051637	55712502166	PAYMENT	C840514538	M1804441305	5885.56	12/06/2019 12:18	0	0
38091815211	36413265030	PAYMENT	C1768242710	M1971783162	5307.88	15/03/2019 10:44	0	0
10395292270	99737753137	PAYMENT	C247113419	M151442075	5031.22	17/07/2019 11:31	0	0
43092321587	41623374421	PAYMENT	C1238616099	M70695990	24213.67	05/03/2019 08:02	0	0
22879832456	38609833651	PAYMENT	C1608633989	M1615617512	8603.42	30/05/2019 11:56	0	0
11117842193	15059372562	PAYMENT	C923341586	M107994825	2791.42	05/04/2019 09:00	0	0

12161951553	57536273892	PAYMENT	C1470868839	M1426725223	7413.54	06/06/2019 10:43	0	0
65298871283	47740065768	PAYMENT	C711197015	M1384454980	3295.19	08/06/2019 14:14	0	0
49062112954	52988074611	PAYMENT	C1481594086	M1569435561	1684.81	20/06/2019 10:05	0	0
95134074779	43896733863	DEBIT	C1466917878	C1297685781	5758.59	21/03/2019 09:18	0	0
30781143535	49434114587	CASH_OUT	C768216420	C1509514333	110414.7	30/03/2019 15:16	0	0
89122906181	17152905095	PAYMENT	C260084831	M267814113	7823.46	09/05/2019 12:22	0	0
62175311757	88392941643	PAYMENT	C598357562	M1593224710	5086.48	06/05/2019 14:13	0	0
17918342582	34177151725	PAYMENT	C1440738283	M1849015357	5281.48	20/03/2019 12:07	0	0
18362821263	31542393502	CASH_OUT	C512549200	C248609774	5346.89	05/07/2019 11:54	0	0

Appendix B

Sample Source Code

B1 – Python

```
#importing modules
import datetime
from os.path import exists
from pathlib import Path
import random as rd
import numpy as np
import pandas as pd
import pickle as pk
#importing modules for visualization
from pandas_profiling import ProfileReport
import matplotlib.pyplot as plt
%matplotlib inline
from matplotlib.colors import ListedColormap
import seaborn as sb
#generating doummy data for bvn (bank verification number) and nin (national identity number)
rd.seed(30)
bvn = [int(str(rd.randrange(1,23121212211,5))[0:7] + str(rd.randrange(1,23121212211,5))[0:4])
for x in range (1480)]
nin = [int(str(rd.randrange(1,59999999999,5))[0:7] + str(rd.randrange(1,59999999999,5))[0:4])
for x in range (1480)]
df_nat = pd.DataFrame({'bvn':bvn,'nin':nin})
df_nat
#df_nat.info()
#creating the first dataframe to get some features
df1 = pd.read_csv(r"C:/Users/HampoJohnPaulAC/MyThesisDataset/ML.csv")
df1
#df1.info()
df1_ = df1[:1480]
#creating the second dataframe to get some features
df2 = pd.read_csv(r"C:/Users/HampoJohnPaulAC/MyThesisDataset/MLtag.csv")
df2
#df2.info()
df2_ = df2[:1480]
#creating the third dataframe to get some features
df3 =
pd.read_csv(r'C:/Users/HampoJohnPaulAC/MyThesisDataset/PS_20174392719_1491204439457_log.csv')
df3
#df3.info()
df3_ = df3[:1480]
#combining all the dataframes into a dataframe
```

```
df = pd.concat([df_nat, df1_, df2_, df3_], axis = 1)
df
#df.info()
df.dtypes
df.columns
df.isna().sum()
df.isnull().sum()
df.dropna()
PR = ProfileReport(df)
PR
# saving the profile report
file_path_profile_report = 'MyThesisDataset/profile_report.html'
file_exist_profile_report = exists(file_path_profile_report)
if file_exist_profile_report == True:
   save_as = 'MyThesisDataset/profile_report.html_' + str(datetime.date.today()).replace("-", "_")
+ '.csv'
   PR.to_file(output_file=save_as)
else:
   save = file_path_profile_report
   PR.to_file(output_file=save)
#saving the combined dataframe
file_path = 'MyThesisDataset/My_Thesis_Sample_dataset.csv'
file_exist = exists(file_path)
if file_exist == True:
   save_as = 'MyThesisDataset/My_Thesis_Sample_dataset_' +
str(datetime.date.today()).replace("-", "_") + '.csv'
   df.to_csv(save_as)
else:
   save = file_path
   df.to_csv(save)
#importing modules for ML
from sklearn import metrics
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
#Detection Model
#assigning and declaring X and y
data = df[['bvn', 'nin', 'typeofaction', 'sourceid', 'destinationid',
      'amountofmoney', 'date', 'isfraud', 'typeoffraud', 'guiltyid',
      'levelofcrime']]
data.columns
X_raw = data[["typeofaction",
      'amountofmoney']]
hamplus = {"typeofaction":     {"cash-in": 0, "transfer": 1}}
X_raw = X_raw.replace(hamplus)
X = X_raw.values
```

```
y = data["isfraud"].values

#spliting data into 75% train and 25% test
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.75, random_state = 30)
#feature scaling
sc = StandardScaler()
X_train_sc = sc.fit_transform(X_train)
X_test_sc = sc.transform(X_test)
#fitting clasifier into the training set
model_classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
detectModel = model_classifier.fit(X_train_sc, y_train)
#saving the model as a pickle file
pk.dump(detectModel,open("MyThesisDataset/detectModel.pkl",'wb'))
#predicting the test set result
y_pred = model_classifier.predict(X_test_sc)
#the confusion matrix
cm = metrics.confusion_matrix(y_test, y_pred)
print(cm)
#visualizing the training set result
X_set, y_set = X_train_sc, y_train
X1, X2 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step =
0.01),
                     np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1,                    X2,                    model_classifier.predict(np.array([X1.ravel(),
X2.ravel()]).T).reshape(X1.shape),
             alpha = 0.25, cmap = ListedColormap(('red', 'green')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
for i, j in enumerate(np.unique(y_set)):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(("red", "green"))(i), label = j)
plt.title("K-NN (Training Set)")
plt.xlabel("Type of Action")
plt.ylabel("Amount of Money")
plt.legend()
plt.show()
#visualizing the test set result
for i, j in enumerate(np.unique(y_set)):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(("red", "green"))(i), label = j)
plt.title("K-NN (Test Set)")
plt.xlabel("Type of Action")
plt.ylabel("Amount of Money")
plt.legend()
plt.show()
#Confusion Matrix of Detection Model
```

```
#['Actual Neg', 'Actual Pos']
sb.heatmap(cm, annot=True, fmt='g', cmap='Blues', yticklabels=['Actual Neg', 'Actual Pos'], xticklabels=['Pred Neg', 'Pred Pos'])
plt.show()
#Prevention Model
#assigning and declaring X and y
data = df[['bvn', 'nin', 'type', 'nameOrig', 'nameDest',
      'amount', 'date', "isFraud", "isflagged"]]
data.columns
X_raw = data[["type",
      'amount']]
hamplus = {"type":        {"PAYMENT": 0, "TRANSFER": 1, 'CASH_OUT': 2, 'DEBIT': 3, 'CASH_IN': 4}}
X_raw = X_raw.replace(hamplus)
X = X_raw.values
y = data["isflagged"].values
#spliting data into 75% train and 25% test
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.75, random_state = 30)
#feature scaling
sc = StandardScaler()
X_train_sc = sc.fit_transform(X_train)
X_test_sc = sc.transform(X_test)
#fitting clasifier into the training set
model_classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
preventModel = model_classifier.fit(X_train_sc, y_train)
#saving the model as a pickle file
pk.dump(preventModel,open("MyThesisDataset/preventModel.pkl",'wb'))
#predicting the test set result
y_pred = model_classifier.predict(X_test_sc)
#the confusion matrix
cm = metrics.confusion_matrix(y_test, y_pred)
print(cm)
#calculating the metrics of Prevention model
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
accuracy = accuracy_score(y_test, y_pred)
print('Accuracy: %.3f' % accuracy)
precision = precision_score(y_test, y_pred, average='binary')
print('Precision: %.3f' % precision)
recall = recall_score(y_test, y_pred, average='binary')
print('Recall: %.3f' % recall)
score = f1_score(y_test, y_pred, average='binary')
print('F-Measure: %.3f' % score)
def specificity_score(y_test, y_pred):
    cm = confusion_matrix(y_test, y_pred)
    return cm[0, 0] / (cm[0, 0] + cm[0, 1])
```

```
specificity = specificity_score(y_test, y_pred)
print('specificity: %.3f' % score)
```

B2 - HTML

```
<!DOCTYPE html>
<html lang="en">

<head>
   <meta charset="UTF-8">
   <meta http-equiv="X-UA-Compatible" content="IE=edge">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <title>MLDPS || Hampo, J.A.C.</title>
   <!-- CSS only -->
            <link      href="https://cdn.jsdelivr.net/npm/bootstrap@5.1.1/dist/css/bootstrap.min.css"
rel="stylesheet"
                                                                                integrity="sha384-
F3w7mX95PdgyTmZZMECAngseQB83DfGTowi0iMjiWaeVhAn4FJkqJByhZMI3AhiU"
crossorigin="anonymous">
      <link rel="shortcut icon" href="{{ url_for('static', filename='images/punishment-for-money-
laundering-in-india.jpg')}}" type="image/x-icon">
      <style>
         p {font-family: 'Times New Roman', Times, serif !important;  font-size: 20px !important;
text-align: justify;}
      </style>
</head>

<body class="bg-dark text-light">
   <div class="container">
      <div class="row mx-auto mt-5">
         <div class="col-3"></div>
         <div class="col-6">
            <h1 class="mb-5 text-center">
               Money Laundering Detection and Prevention System
            </h1>
         </div>
         <div class="col-3"></div>
      </div>
      <div class="row">
         <div class="col-3"></div>
         <div class="col-6">
            <nav class="nav nav-bar">
               <div class="col-2">
                        <a href="{{ url_for ('home') }}" style="pointer-events: none;"><button
type="submit" class="btn btn-light mb-3">Home</button></a>
               </div>
```

```
            <div class="col-2">
               <a href="{{ url_for ('login') }}"><button type="submit"
                     class="btn btn-success mb-3">Login</button></a>
            </div>

            <div class="col-2">
               <a href="{{ url_for ('about') }}"><button type="submit"
                     class="btn btn-warning mb-3">About</button></a>
            </div>

            <div class="col-3">
               <a href="{{ url_for ('developer') }}"><button type="submit"
                     class="btn btn-danger mb-3">Developer</button></a>
            </div>

            <div class="col-2">
               <a href="{{ url_for ('contribute') }}"><button type="submit"
                     class="btn btn-primary mb-3">Contribute</button></a>
            </div>
         </nav>
      </div>
      <div class="col-3"></div>
   </div>

   {% block body %}


   {% endblock %}

  </div>
</body>

</html>

<!DOCTYPE html>
<html lang="en">

<head>
  <meta charset="UTF-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>MLDPS || Hampo, J.A.C.</title>
  <!-- CSS only -->
        <link      href="https://cdn.jsdelivr.net/npm/bootstrap@5.1.1/dist/css/bootstrap.min.css"
rel="stylesheet"
```

```
                                                                    integrity="sha384-
F3w7mX95PdgyTmZZMECAngseQB83DfGTowi0iMjiWaeVhAn4FJkqJByhZMI3AhiU"
crossorigin="anonymous">
      <link rel="shortcut icon" href="{{ url_for('static', filename='images/punishment-for-money-
laundering-in-india.jpg')}}" type="image/x-icon">
</head>

<body class="bg-dark text-light">
   <div class="container">
      <div class="row">
         <div class="col-12 mx-auto mt-3 text-center">
            <h1 class="mb-3">PREVENTION</h1>
         </div>
      </div>
      <form action="{{ url_for ('prevent') }}" method="post">
         <div class="row">
            <div class="col-6">
               <div class="input-group mb-3">
                  <span class="input-group-text" id="basic-addon3">NIN</span>
                   <input type="text" class="form-control" id="basic-url" aria-describedby="basic-
addon3" name="nin" maxlength="11" required>
               </div>
            </div>

            <div class="col-6">
               <div class="input-group mb-3">
                  <span class="input-group-text" id="basic-addon3">BVN</span>
                   <input type="text" class="form-control" id="basic-url" aria-describedby="basic-
addon3" name="bvn" maxlength="11" required>
               </div>
            </div>

            <div class="col-6">
            <fieldset class="border p-2 mb-2">
               <legend>Source Account Details</legend>
               <div class="input-group mb-3">
                  <span class="input-group-text" id="basic-addon3">Account Number</span>
                   <input type="text" class="form-control" id="basic-url" aria-describedby="basic-
addon3" name="accountNumberS" maxlength="10" required>
               </div>

               <div class="input-group mb-3">
                        <select class="form-select" size="4" aria-label="size 4 select example"
name="accountTypeS" required>
                     <option selected>Account Type</option>
                     <option value="1">Tier-1</option>
```

```
                    <option value="2">Tier-2</option>
                    <option value="3">Tier-3</option>
                  </select>
                </div>

                <div class="input-group mb-3">
                  <span class="input-group-text" id="basic-addon3">Source ID</span>
                   <input type="text" class="form-control" id="basic-url" aria-describedby="basic-addon3" name="sourceID" maxlength="6" title="Sending Bank Sort Code" placeholder="Sending Bank Sort Code" required>
                </div>
              </fieldset>
              </div>

              <div class="col-6">
              <fieldset class="border p-2 mb-2">
                <legend>Destination Account Details</legend>
                <div class="input-group mb-3">
                  <span class="input-group-text" id="basic-addon3">Account Number</span>
                   <input type="text" class="form-control" id="basic-url" aria-describedby="basic-addon3" name="accountNumberD" maxlength="10" required>
                </div>

                <div class="input-group mb-3">
                        <select class="form-select" size="4" aria-label="size 4 select example" name="accountTypeD" required>
                    <option selected>Account Type</option>
                    <option value="1">Tier-1</option>
                    <option value="2">Tier-2</option>
                    <option value="3">Tier-3</option>
                  </select>
                </div>

                <div class="input-group mb-3">
                <span class="input-group-text" id="basic-addon3">Destination ID</span>
                   <input type="text" class="form-control" id="basic-url" aria-describedby="basic-addon3" name="destinationID" maxlength="6" title="Receiving Bank Sort Code" placeholder="Receiving Bank Sort Code" required>
              </div>
              </fieldset>
            </div>

            <div class="col-6">
              <div class="input-group mb-3">
                <span class="input-group-text" id="basic-addon3">Transaction ID</span>
```

```
                <input type="text" class="form-control" id="basic-url" aria-describedby="basic-addon3" name="transactionID" value="{{ transactionID }}" readonly required>
            </div>
            </div>

            <div class="col-6">
            <div class="input-group mb-3">
                        <select class="form-select" size="3" aria-label="size 4 select example" name="transactionType" required>
                  <option selected>Transaction Type</option>
                  <option value="1">Cash-in</option>
                  <option value="2">Cash-out</option>
                </select>
            </div>
          </div>

          <div class="col-6">
            <div class="input-group mb-3">
              <span class="input-group-text">Amount (&#8358)</span>
                <input type="text" class="form-control" aria-label="Amount (to the nearest naira)" name="amount" required>
              <span class="input-group-text">.00</span>
            </div>
            </div>

            <div class="col-6">
            <div class="input-group mb-3">
              <span class="input-group-text">Previous Bal (&#8358)</span>
                <input type="text" class="form-control" aria-label="Amount (to the nearest naira)" name="previousBalance" required>
              <span class="input-group-text">.00</span>
            </div>
            </div>

            <div class="form-check form-switch" hidden>
              <input class="form-check-input" type="checkbox" id="flexSwitchCheckDefault">
              <label class="form-check-label" for="flexSwitchCheckDefault">IsFlagged</label>
            </div>

            <div class="form-check form-switch" hidden>
                <input class="form-check-input" type="checkbox" id="flexSwitchCheckDefault" value="1" name="isFlagged">
              <label class="form-check-label" for="flexSwitchCheckDefault">IsFraud</label>
            </div>

            <div class="col-auto">
```

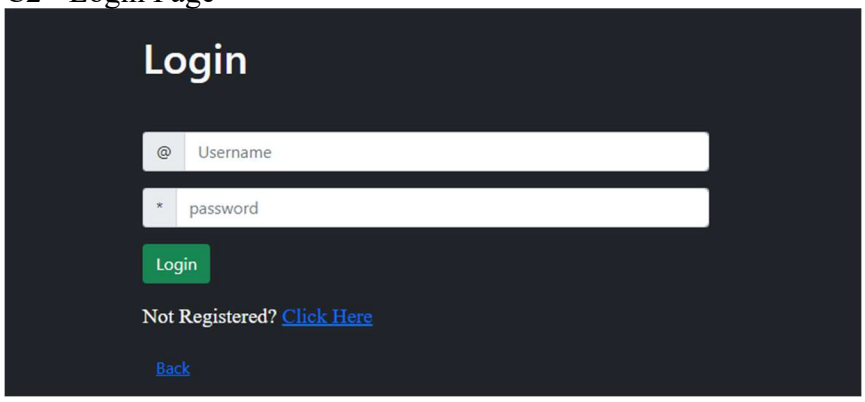
```
                                    <button type="submit" class="btn btn-success mb-3"
name="detect">Prevent</button>
            </div>

            <div class="col-auto">
            <a href="{{ url_for ('dashboard') }}">
               <button type="button" class="btn btn-primary text-light">Back</button>
            </a>
          </div>
          </form>
       </div>
    </div>
</body>
</html>
```

Appendix C

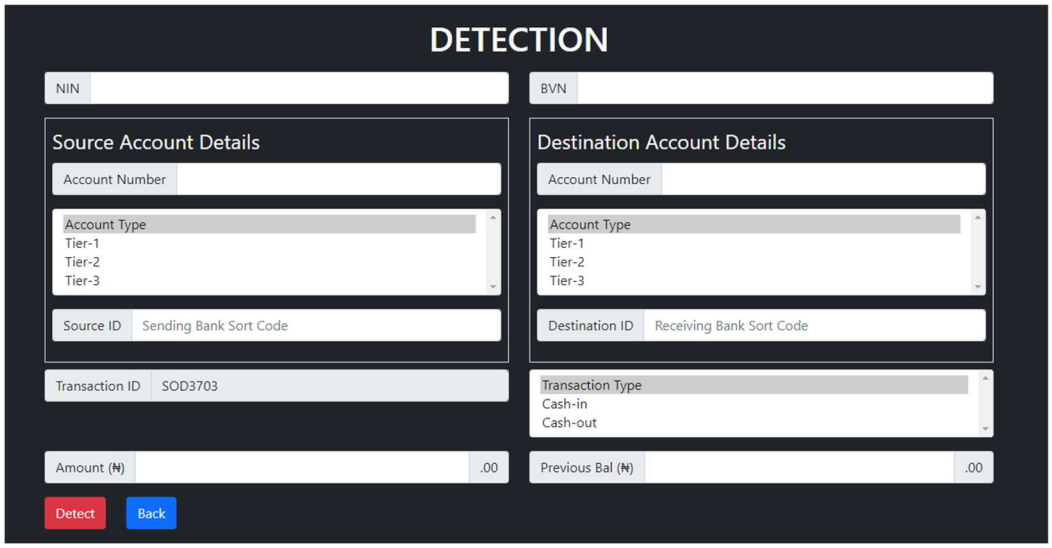
Sample Interfaces

C1 - Home Page

C2 - Login Page



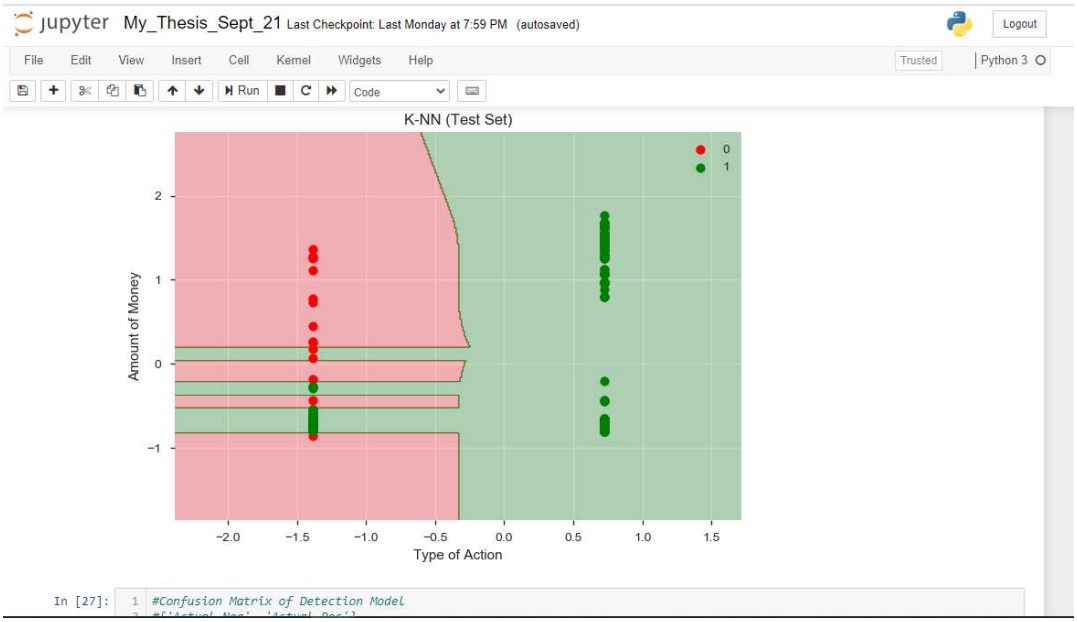
C3 - Detection Page



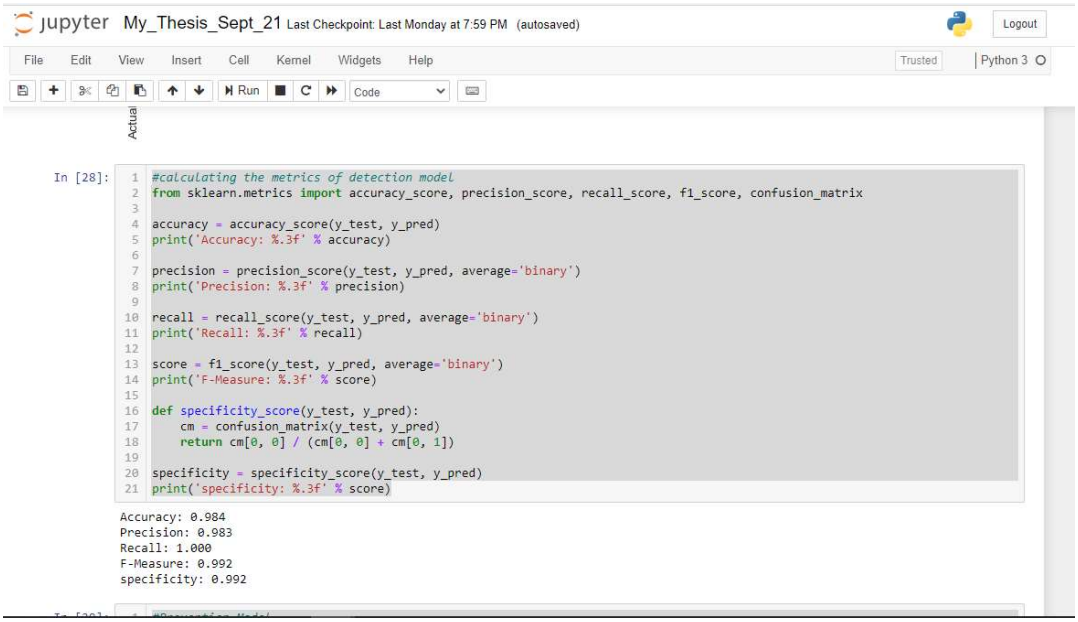
Appendix D

Results

D1 – Jupyter IDE Showing Visualization



D2 – Jupyter IDE Showing Metrics Code



Appendix E

Questionnaire

Personal Interview and Questionnaire Survey

Dear Sir/Ma,

I am a graduate of Computer Science from Delta State University, Abraka, currently undertaking a Master's Degree program in Computer Science with Federal University of Technology, Owerri. My supervisors are Dr. (Mrs) E.C. Nwokorie and Dr. (Mrs) J.N. Odii.

This survey/interview is to enable the collection of information on money laundering, which will be used to develop an anti-money laundering solution using machine learning and social network analysis algorithms. I, therefore, beseech you to provide unprejudiced answers to the questions herein.

Your time taken for this survey/interview is highly appreciated, as it will take 5 minutes or at most 10 minutes to complete the survey/interview. Responses are anonymous and no sensitive data will be thrown to the wind.

You are free to withdraw from the survey at any time you wish. Your decision to withdraw will have absolutely no repercussions. Your consent to participate in this survey should be shown by signing _____.

Your participation in this survey/interview has no known risk or benefit. However, the researcher hopes that the data gathered will provide information about the link between money laundering and developing an anti-money laundering solution.

If you have any questions or require any form of clarification concerning this survey, please contact the researcher at***:*** ***hampojohnpaul@gmail.com*****,*** ***founder@hamplustech.com*****,*** ***+2348050736053,*** ***+2347063047037***

Thank you.

Hampo, JohnPaul A.C.

Researcher

A Web Based Money Laundering Detection and Prevention Model Using Machine Learning and Social Network Analysis

Section A: Role in bank is either customer service branch manager, bank branch manager, fraud control unit branch manager

Gender: Male ( ) Female ( ) *[Please tick one]*

Role in Bank: _____

Experience in Role: _____

Section B:

1. Is money laundering detection in Nigeria automated? Yes ( ) No ( )
2. Please name the software(s) used in anti-money laundering.

_____

3. What are the inadequacies in the existing anti-money laundering system?

_____

_____

_____

4. Would you like an improvement on the identified inadequacies? Yes ( ) No ( )
5. Please specify the channel(s) or medium(s) through which money transfer is done in Nigeria

_____

_____

_____

6. Does all money transfer in Nigeria passes through ONE gateway, irrespective of the source channel, medium, platform or app? ____ If yes, please state the gateway _____
7. Please specify the different account types and their maximum single deposit limit in Nigeria

_____

_____

_____

8. Please specify the different account types and the maximum daily transaction limit in Nigeria

_____

_____

_____

9. Please specify the different account types and the maximum balance limit in Nigeria

_____

_____

_____

10. In Nigeria, is the maximum balance in an account dependent on the account type or the individual(s)? Yes ( ) No ( )
11. What is the maximum amount a BVN can hold? _____

_____

12. Is money transfer from outside Nigeria subject to the financial limit placed on account types in Nigeria? Yes ( ) No ( )
13. Is money transfer from outside Nigeria subject to the financial limit placed on BVN in Nigeria? Yes ( ) No ( )

14. Briefly state what can make an account to be flagged as suspicious in money laundering in Nigeria

_____

_____

_____