BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

By

NYAMUTSAMBA CHIPO TALENT

B192394B

SUPERVISOR: MR MUZURURA

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE BACHELOR OF SCIENCE HONOURS DEGREE IN COMPUTER SCIENCE

APPROVAL FORM

The undersigned certify that they have supervised the student Nyamutsamba Chipo Talent's dissertation entitled, "Credit Card Fraud Detection Using Machine Learning" submitted in partial fulfilment of the requirements for a Bachelor of Computer Science Honours Degree at Bindura University of Science Education.

STUDENT: DATE:

…… ……..

SUPERVISOR: DATE:

…… .…….

CHAIRPERSON: DATE:

…… .…….

EXTERNAL EXAMINER: DATE:

…… .…….

ABSTRACT

Fraudsters steal a lot of money every day, so it's critical to develop algorithms that can help cut down on these losses. However, due to the non-stationary distribution of the data, the extremely imbalanced class distributions, and the availability of few transactions tagged by fraud investigators, the construction of fraud detection algorithms is particularly difficult. The optimal course of action is unclear because there is a lack of publicly available information due to confidentiality concerns. To produce reliable alerts, detection systems must be able to handle the demands of real-world operations and incorporate feedback from researchers. The most widely used payment methods both online and offline are credit cards. The increased use of credit cards globally is also leading to an increase in fraud.

Every store in the world accepts credit cards for cashless transactions. A fraudulent transaction will be discovered after it has been completed by the credit card fraud detection system. To identify, examine, and stop credit card fraud, several methods have been created. Examples of these include machine learning algorithms and techniques, big data analytics, and artificial neural networks. The objective of this study is to forecast the incidence of fraud utilizing machine learning methods and algorithms including Logistic Regression.

KEYWORDS: Big Data Analytics, Artificial Neural Networks, Machine Learning

DECLARATION

I NUNURAI PHINIAS CHITENGU, hereby declare that this project is my original work and I am the sole author of this dissertation. I authorize the BINDURA UNIVERSITY OF SCIENCE EDUCATION to lend this dissertation to other institutions and/or relevant firms for the purpose of scholarly research.

Signature ……

Date ……

ACKNOWLEDGEMENT

First and foremost, all the credit goes and belongs to the Almighty God who guided me through my final year dissertation. Mr O Muzurura my supervisor, I extend my sincere gratitude for the time, patience and resilience he invested in me as I worked through my research project. I value your endless time and contribution Sir. I also want to thank Bindura University of Science Education for all the infrastructural support and all the academic provisions that led to the completion of this study. Not forgetting my family and fellow colleagues Trish N Paroreya, Panashe K Gazera, Zivanayi Zvarevashe, and Kimberly Chiguma etc. for playing a supportive role that aided my well-being.

DEDICATION

I want to dedicate this project to my mother, my siblings and my family for their nurturing and continual support they gave me from the beginning of my time up to this moment. Patience, endurance, persistence and passion are fundamental they gave which have been taking me to different levels in life. My right doings are all rooted to these special people.

TABLE OF CONTENTS

1.0 INTRODUCTION

As a result of technological improvements, the emergence of E-commerce, and its acceleration, credit card usage has significantly expanded in recent years. A credit card is a payment card that is given to customers so they can pay a merchant for products and services based on the debt they have accumulated. It provides the cardholder with the benefit of time to repay at a later date. As credit card transactions have been the most frequent mode of payment in recent years, they have given rise to both the positive and negative aspects of utilizing them as a payment method. Credit card theft occurs in all industries nowadays, including the appliances sector, the automobile industry, banks, daily transactions, and so on. Credit card fraud is defined as the unauthorized use of any system or criminal conduct using physical cards or card information without the cardholder's knowledge.

Financial crime, particularly credit card theft, is on the rise in Zimbabwe and around the world, boosted by technology improvements. Banks face technical hurdles, a lack of information and awareness, and a lack of legislation. Banks lose goodwill and customer trust, and businesses lose credibility as a result of credit card fraud. The financial losses caused by banks, enterprises, and cardholders are detrimental to users' activity.

Credit card fraud detection can be improved by using machine learning techniques. Algorithms can distinguish between genuine and fraudulent transactions. The algorithms will categorise all transactions by working with a set of datasets.

1.1 BACKGROUND TO THE AREA OF STUDY

Since fraudulent use of credit cards has become such a large concern in the financial industry, several financial institutions have spent significant money and formed teams of human professionals to build fraud detection systems. Many academics have worked hard to minimize various obstacles that arise while creating a fraud detection system (FDS), such as class imbalance, class overlapping, changes in fraud behaviours, and so on.

A report published by the Zimbabwe Republic Police in March disclosed that in the first quarter of 2018, at least $200,000 had been lost to bank card cloning. In addition, the year 2020 on the 16th day of August, it was reported that Nyaradzo lost $200K in suspected case of card cloning.

Furthermore, the following year that is 2021 had two reported cases about credit card fraud. On the first day of January, a man was arrested for stealing $500K in bank card fraud and on 18 March, a man was reported to have cloned bank cards and stealing Z$3Million.

1.2 PROBLEM STATEMENT

The number one business priority for many banks is to maintain highly profitable customers. Banking fraud, on the other hand, offers a huge threat to this goal for various institutions in terms of major financial losses, confidence and credibility; this is a worry for both banks and customers.

Credit card fraud detection employing machine learning is not only an expectation in the banking industry, but it is also an essential for them to have proactive monitoring and fraud protection measures. Machine learning is assisting these institutions in reducing time-consuming manual checks, costly charge backs and fees, and legitimate transaction denials.

The main focus of this work is on how to detect credit card fraud using machine learning algorithms, given the datasets, and categorize which transactions are credit card fraud or not.

1.3 OBJECTIVES

The researcher seeks to achieve the following objectives on the research project:

- To examine on the various techniques of credit card fraud detection and taking a look at their weaknesses.
- To develop a system that correctly classify transactions into their categories, either fraudulent or legitimate, in real-time and with adaptiveness.
- To test, analyse and evaluate the developed credit card fraud detection.

1.4 RESEARCH QUESTIONS

The researcher will seek to answer the following questions on the research:

- What “on-line” credit card fraud detection methodologies are available and what are their weaknesses?
- How best can we improve the current systems to come up with a system that can correctly classify online- transactions?
- How can we measure the accuracy and error rate of the developed system?

1.5 RESEARCH HYPOTHESIS

- H_0: Machine Learning techniques result in optimized credit card fraud detection.

- H_1: Machine Learning techniques will not result in optimized credit card fraud detection

1.6 SIGNIFICANCE OF THE STUDY

Credit card fraud detection relates to the standards, approaches, methodologies, as well as procedures used by credit card issuers and financial institutions to combat and stop fraudulent transactions.

With this level of control, fraudsters lack the ability to perform multiple transactions on a stolen or cloned card before the cardholder becomes aware of the fraudulent activity. This can save a significant sum of money that would otherwise be lost due to fraud.

1.7 SCOPE OF THE STUDY

The study focuses on Machine Learning techniques to develop an A.I system that will be able to effectively detect fraudulent credit card transactions in a given dataset. The research is intended for the cardholders, financial and banking industry, helping also Data Scientists and Analysts in analysing the data given by the findings

.

1.8 LIMITATIONS AND CHALLENGES

- There were challenges to datasets access.
- The system could not be executed on an authentic platform.
- There was only a short period of time spent gathering data and identify trends.

1.9 DEFINITION OF TERMS

Big Data – big data refers to sets of data that are both vast in size and intricate in complexity. They are so immense that conventional data processing methods and tools are insufficient for analysis. These data sets can consist of structured, semi-structured, or unstructured data and originate from various sources such as social media, weblogs, sensor data, and scientific research, among others.

Big Data Analytics – Big Data Analytics refers to the analysis of vast and intricate data sets that cannot be processed or analysed using conventional data processing tools and methods. This involves utilizing specialized techniques and software to extract valuable insights and meaning from semi-structured or unstructured data.

Artificial Neural Networks – Artificial Neural Networks (ANNs) are a type of machine

learning algorithm that imitates the way the human brain operates. They are composed of interconnected nodes, referred to as neurons that carry out computations and transmit information through multiple layers.

Machine Learning – Machine Learning (ML) is a branch of computer science that concentrates on creating algorithms capable of analysing data and making predictions or decisions based on that data. It is a subset of artificial intelligence that involves building models or programs that can "learn" from data without human intervention or explicit programming.

CHAPTER TWO: LITERATURE REVIEW

2.1 INTRODUCTION

The purpose of examining relevant literature is to help the researcher align their current investigation with previous research studies on specific issues related to their study. The aim is to evaluate what research has already been conducted, what is currently in progress, and what needs to be accomplished in the research project.

(Borg and Gall, 1979) blend the above statement when they acknowledge that "…according to the ethics of social science, one should try to read the latest material

because continuous research is continuously bringing out new information.”

The internet and the growth of E-commerce have led to a significant increase in the use of credit cards, which has also resulted in a rise in credit card fraud. Given the prevalence of credit card theft in the digital age, it is essential to detect and minimize such fraudulent activities. Financial fraud has significant impacts on both the financial industry and individuals' daily lives, including businesses, financial institutions, and customers. To address this issue, many researchers are using Machine Learning, a field of Artificial Intelligence, to tackle various problems.

The primary objective of this section is to explore different approaches to credit card fraud detection. The focus is on presenting, analysing, and comparing previous research findings in credit card fraud detection, highlighting their differences, limitations, and contributions. This is intended to identify gaps in the existing literature and to explain how the current study aims to address these gaps.

2.2 RELEVANT THEORIES ON THE SUBJECT MATTER

Credit enables the sale of goods or services to buyers who do not have cash at hand. A credit card is a small, thin plastic or fibre card that contains personal information, such as a picture or signature, and is used to charge purchases and services to the individual's account, which is deducted regularly. Credit card information can be read by various devices, such as ATMs, retail readers, swipe machines, banks, and online transactions. The card's physical security and privacy are critical elements of its security.

(Oxford English Dictionary, 2021) defined fraud as “wrongful or criminal deception intended to result in financial or personal gain". Credit card fraud is also defined as the unlawful and unwanted use of an account by an individual not related to the account owner or the card-issuing authorities who are unaware of its use. Measures can be implemented to prevent instances of fraud and safeguard against potential future occurrences. However, since relying on human detection of fraud is not entirely reliable, credit card fraud must be approached with caution by utilizing machine learning.

Fraud detection comprises watching the actions of user populations in order to estimate, recognize, or avoid undesired behaviour such as fraud, intrusion, and defaulting. It comprises monitoring and analysing various users' patterns of conduct in order to estimate, recognize, and avoid undesirable behaviour. Methods of detecting fraud are always being developed in order to prevent criminals from adapting to detection strategies.

Developing new strategies for credit card fraud detection is a time-consuming and challenging process, which can cause delays in obtaining new results. Financial institutions, such as banks, are often hesitant to provide data for use in fraud detection datasets. As a result, platforms such as Google and Kaggle are used to supply datasets for this purpose. With the aid of vast amounts of data, various fraud detection techniques, including artificial intelligence, data mining, and statistics, can be employed to predict instances of fraud.

Distinguishing between legitimate and fraudulent transactions can be challenging, especially when relying on basic pattern matching techniques. To minimize losses, credit card issuers and banks utilize various technologies for credit card fraud detection. These technologies include Artificial Intelligence, Data Mining, Neural Networks, Bayesian Networks, Fuzzy Logic, Artificial Immune Systems, K-Nearest Neighbor Algorithm, Support Vector Machines, Decision Trees, Fuzzy Logic-Based Systems, Machine Learning, Sequence Alignment, Genetic Programming, and more.

Fraud can take on various forms, including theft fraud and counterfeit fraud, which are closely related. Theft fraud involves the unauthorized usage of a credit card by someone other than the owner. In such cases, the owner should report the incident to

the bank or credit card issuer, who will take appropriate measures. Counterfeit fraud, on the other hand, involves the use of credit card information remotely, without the need for the physical card itself.

This chapter discusses several types of financial fraud, including telecommunication fraud, computer intrusion, application fraud, and bankruptcy fraud.

Telecommunication fraud, also referred to as telco fraud or telecom fraud, is a type of fraud where a perpetrator utilizes telecommunication services to commit other forms of fraud, such as identity theft or financial fraud. This type of fraud can impact businesses and communication services.

Computer intrusion refers to the unauthorized entry or attempted entry into a computer system, without proper authorization or invitation. This means that there is a potential for unauthorized access to information or manipulation of information with malicious intent. Intruders can come from any environment, including outside hackers or insiders who have knowledge of the system's layout.

Bankruptcy fraud can manifest in various ways. One form of bankruptcy fraud involves using a credit card while being insolvent, meaning that consumers knowingly use credit cards despite being unable to pay for their purchases, thereby falsifying information. Another form of bankruptcy fraud involves the use of a credit card while the owner is absent. This type of fraud is complex and difficult to detect.

Application fraud takes place when an individual applies for a credit card from a bank or credit card issuer using false information. To detect application fraud, two types of situations need to be identified. The first is duplicates, which occur when applications come from the same user with the same details. The second type is identity fraud, which occurs when applications come from different individuals with similar details. Phua et al. (2006) Defined application fraud as "the use of false or misleading information by individuals or organizations in order to obtain something of value, such as a loan, a credit card, or a job, to which they would not otherwise be entitled." This form of fraud can occur in various contexts, such as credit card applications, mortgage applications, and loan applications.

Credit card fraud can be divided into two types thus Offline fraud and On-line fraud.
Online credit card fraud is a type of online fraud that involves the unauthorized use of a credit card for fraudulent transactions through online channels such as e-commerce websites, mobile apps, and payment gateways. This makes it possible for criminals to intercept your data and use it to make unauthorized purchases.
Offline credit card fraud refers to the unlawful use of a credit card for fraudulent transactions that take place in the physical world, outside of the online realm. This type of financial fraud can involve different methods, such as the use of lost or stolen credit cards, the installation of skimming devices, and the creation of fake credit cards.

2.2 CHARACTERISTICS OF CREDIT CARD FRAUD DETECTION SYSTEMS

Fraud detection is a difficult undertaking, and detection systems are prone to a range of hurdles and constraints. The system must evaluate both the cost of fraudulent

behaviour and the cost of preventing it from occurring. A decent credit card fraud detection system should be capable of the following:

- Identify the frauds accurately.
- Detect the frauds quickly.
- Distinguish and categorize legitimate transactions from fraudulent transactions.

2.2.1 CHALLENGES THAT ARE LIKELY TO ARISE IN DETECTING CREDIT CARD FRAUD

The advent of the credit card issue introduces additional challenges to the detection process. Some of the difficulties that are likely to arise are as follows:

i) **Changing fraud patterns over time** – Identifying and managing this type of fraud is challenging because fraudsters are continuously seeking new and innovative ways to evade detection systems. To prevent such fraud, deep learning models need to be regularly updated with the latest detection techniques; otherwise, machine learning models will not be able to achieve their objectives.

ii) **Class imbalance –** A small fraction of customers have fraudulent intentions, creating an imbalance in fraud detection models that typically classify transactions as either fraudulent or non-fraudulent. This imbalance poses a challenge in building such models. As a consequence, genuine customers may have a negative user experience because detecting fraudsters may require the rejection of valid transactions.

iii) **Model interpretations –** The limitation of explain ability in models is associated with the fact that they provide a score indicating the likelihood of a transaction being fraudulent or not, without offering any explanation as to why.

iv) **Feature generation can be time consuming –** The fraud detection process may be slowed down as subject matter experts take a considerable amount of time to develop a comprehensive set of features.

2.2.2 APPROCHES TO ADDRESSING CREDIT CARD FRAUD DETECTION

The difficulties can be addressed using a variety of accessible solutions. Some of them are:

i. **Ensemble Modelling** – Ensemble modeling is a technique that involves using multiple models to perform a single task, such as identifying instances of fraud. This approach is useful for detecting fraudulent behavior that continuously evolves and changes over time. By combining different types of models, such as classical machine learning, deep learning, and linear models, it is possible to capture a wide range of fraudulent behaviors and maximize the accuracy of fraud detection. For instance, a deep learning model like LSTM can effectively identify fraud in a sequence of events. If a user logs in from an unusual location, makes changes to their address book, and then makes a high-cost purchase on an e-commerce site, the LSTM model may flag the transaction as fraudulent.
ii. **Commerce use case** – Extrapolation of a model information and application to various use cases such as when people relocate email addresses rather than physical addresses. The program learns from the examples presented and then identifies new ones depending on human input
iii. **Explain ability** - The concept of explainable AI involves providing explanations for

why a particular transaction is either approved or rejected as fraudulent. This approach can help to overcome the challenge of interpreting models used in fraud detection. Various techniques can be used to achieve explain ability, such as surrogate modelling, maximal activation analysis, and others. These techniques can provide specific benefits in terms of explaining the reasoning behind the model's decisions.

2.3 CONCEPTUAL FRAMEWORK

MACHINE LEARNING AND DETECTION TECHNIQUES

Machine Learning is a branch of artificial intelligence that enables computers to learn from experience without the need for human intervention. It involves the use of various algorithmic models to accurately predict future outcomes. It is also a type of artificial intelligence technique that utilizes algorithms to enable robots to learn from data. Machine Learning focuses on the study of pattern recognition and computational learning theory within the field of artificial intelligence.

Supervised, Unsupervised, and Reinforcement learning are the three types of Machine Learning algorithms used to detect credit card fraud. It is determined by the nature of the "signal" or "feedback" available to the learning system. They are as follows:

- SUPERVISED LEARNING

Supervised Machine Learning is an algorithm that learns from labeled training data to help you predict outcomes for unforeseen data. In Supervised learning, you train the machine using data that is well “labeled.” It means some data is already tagged with correct answers.

Supervised machine learning involves training a model using both input and output labels. These labels are provided by a "teacher" who gives the computer example inputs and their corresponding desired outputs, in order to teach the model a general rule that maps inputs to outputs. The model then uses this rule to generate prediction models to solve related problems. A label in this context refers to a classification, which involves categorizing data into one or more groups.

SUPERVISED LEARNING TECHNIQUES

The following are various supervised learning techniques

1. **Linear regression:** A technique used for predicting a continuous output variable based on one or more input variables.
2. **Logistic regression:** A technique used for binary classification problems, where the goal is to predict one of two possible outcomes based on input variables.
3. **Decision trees:** A technique used for both classification and regression problems, where the model is represented as a tree-like structure that partitions the data based on input variables.
4. **Random forests:** An ensemble technique that combines multiple decision trees to improve accuracy and reduce overfitting.

- UNSUPERVISED LEARNING

Unsupervised learning is a machine learning approach that requires no prior knowledge or supervision, where a model is trained on unlabelled data to discover patterns and relationships within the data. This type of algorithm is used to identify similarities or differences between data points, group them into clusters, and extract features.

UNSUPERVISED LEARNING TECHNIQUES

The various unsupervised learning techniques include

1. **Clustering:** A technique used to group similar data points together into clusters, based on the similarity of their features.
2. **Dimensionality reduction:** A technique used to reduce the number of features in the data, while preserving the most important information.
3. **Anomaly detection:** A technique used to identify unusual or abnormal data points, which may indicate errors or fraud.
4. **Association rule mining:** A technique used to discover relationships between different items in a dataset, such as frequently co-occurring items in a shopping cart.

5. **Neural networks:** In some cases, neural networks can also be used for unsupervised learning, such as in the case of auto encoders, which are used for dimensionality reduction.

- SEMI-SUPERVISED LEARNING

Semi-supervised machine learning is a type of machine learning that combines both labelled and unlabelled data to improve the performance of a model. In semi-supervised learning, the model is trained on a small amount of labelled data and a larger amount of unlabelled data. The labelled data is used to guide the learning process, while the unlabelled data is used to discover patterns and relationships in the data.

SEMI-SUPERVISED LEARNING TECHNIQUES

There are various semi-supervised learning algorithms, including:

1. **Self-training:** In self-training, the model is trained on the labelled data, and then the model is used to make predictions on the unlabelled data. The predictions with the highest confidence are then added to the labelled dataset and used to retrain the model.
2. **Co-training:** In co-training, multiple models are trained on different subsets of the features of the data. The models then make predictions on the unlabelled data,

and the predictions with the highest agreement between the models are added to the labelled dataset and used to retrain the models.

3. **Multi-view learning:** In multi-view learning, multiple models are trained on different views of the data, such as different types of features or different representations of the data. The models are then combined to make predictions on the unlabelled data.

2.4 EMPIRICAL REVIEW

Scientists all across the world are interested in fraud detection, especially credit card fraud detection. Several methods to recognize credit card fraud have been produced depending on clustering techniques, data mining, genetic algorithms, neural system, decision tree, Bayesian systems, Linear regression, and so forth.

Ramya (2019) proposed a machine learning approach using Bayesian and Neural Networks to automatically distinguish credit card fraud detection systems. In their

study, they used a dataset of credit card transactions and applied feature selection to identify the most relevant features for fraud detection. They then used Bayesian Networks and Neural Networks to train models on the selected features and compare their performance.

The results of their study showed that both Bayesian Networks and Neural Networks were effective in detecting credit card fraud, with Neural Networks outperforming Bayesian Networks in terms of accuracy.

Rathore (2016) suggested a hybrid technique for credit card fraud detection. The proposed technique combines rule-based and machine learning-based approaches to improve the accuracy of fraud detection.

The rule-based approach involves defining a set of rules to identify potential fraud cases based on specific features of credit card transactions, such as the transaction amount, location, and time. The machine learning-based approach involves training a model on a labelled dataset of credit card transactions, using features such as transaction amount, merchant category code, and time of transaction.

The proposed hybrid technique uses a decision tree algorithm to combine the results of the rule-based and machine learning-based approaches. The decision tree algorithm evaluates the rules and machine learning models in a hierarchical manner, assigning weights to each approach based on their performance. The final decision is based on the combined results of the rule-based and machine learning-based approaches.

Khandare (2016) proposed the use of a Hidden Markov Model (HMM) to facilitate the prevention of online and offline credit card fraud. The proposed model uses HMM to analyse the sequence of transactions made by a credit card user, and detect any anomalies or deviations from their usual behaviour.

The HMM model is trained on a dataset of the user's transaction history, and learns to identify patterns and relationships between the user's transactions. The trained model is then used to predict the likelihood of a new transaction being fraudulent, based on the observed sequence of transactions. If the model predicts a high likelihood

of fraud, the transaction can be blocked or flagged for further investigation.

The proposed HMM model improves upon traditional rule-based approaches to fraud detection by taking into account the sequence of transactions and the user's behaviour over time. The model can adapt to changes in the user's behaviour and adjust the predictions accordingly.

Mubalaik (2017) proposed the use of an Artificial Neural Network - Multilayer Perceptron (ANN-MPL) for credit card fraud detection. The proposed model uses a feedforward neural network with multiple hidden layers to learn the patterns and relationships between the features of credit card transactions.

The ANN-MPL model is trained on a labelled dataset of credit card transactions, where the target variable indicates whether the transaction is fraudulent or not. The model learns to identify the most important features of the transactions that are most indicative of fraud, and uses this information to make predictions on new transactions.

The proposed ANN-MPL model improves upon traditional machine learning algorithms by allowing for more complex relationships between the features of the transactions. The model can also adapt to changes in the data and learn from new examples, improving its accuracy over time.

Saleem (2022) proposed a deep learning-based credit card fraud detection system that uses auto encoders and convolutional neural networks (CNNs). The system consists of three main components: data pre-processing, feature extraction, and fraud detection.

Overall, the proposed system offers a novel approach to credit card fraud detection using deep learning techniques. The use of auto encoders and CNNs for feature extraction and pattern recognition allows the system to identify subtle patterns in the data that may be indicative of fraud. The authors suggest that their system could be used by credit card issuers and payment processors to improve the accuracy and efficiency of fraud

Zhang (2022) proposed a hybrid credit card fraud detection system that combines gradient boosting decision tree (GBDT) and deep neural network (DNN) algorithms. The system consists of four main stages: data pre-processing, feature engineering, hybrid model construction, and model evaluation.

Overall, the results suggest that the proposed hybrid credit card fraud detection system is effective at detecting fraudulent transactions with high accuracy and recall rates, and has the potential to improve the efficiency and accuracy of fraud detection systems used by credit card issuers and payment processors.

Li (2021) proposed a credit card fraud detection system based on random forest and an improved gradient boosting decision tree (GBDT) algorithm. The system consists of three main stages: data pre-processing, feature selection, and fraud detection.

The results of the study suggest that the proposed credit card fraud detection system based on random forest and improved GBDT algorithms is effective at detecting fraudulent transactions with high accuracy and recall rates. This system has the potential to be useful for financial institutions and payment processors in detecting fraudulent activities. However, further research is needed to evaluate the system's performance on larger datasets and to test its scalability and efficiency in real-world applications.

Singh (2021) proposed a novel credit card fraud detection system that combines machine learning and deep learning techniques. The system consists of four main stages: data pre-processing, feature engineering, model construction, and model evaluation.

The results of the study suggest that the proposed credit card fraud detection system using machine learning and deep learning techniques is effective at detecting fraudulent transactions with high accuracy and recall rates. However, further research is needed to evaluate the system's performance on larger datasets and to test its scalability and efficiency in real-world applications.

Chen et al. (2021) proposed an anomaly detection based credit card fraud detection system using an improved generative adversarial network (GAN). The proposed system consists of three main stages: data pre-processing, feature extraction, and anomaly detection.

The results of the study suggest that the proposed system has the potential to improve credit card fraud detection and reduce the financial losses and risks associated with fraudulent activities.

Ganatra (2021) proposed a credit card fraud detection system using machine learning and deep learning techniques. The system consists of three main stages: data pre-processing, feature engineering, and model construction.

Overall, the proposed credit card fraud detection system using machine learning and deep learning techniques represents a promising approach to fraud detection. The use of a hybrid model allows the system to extract and process relevant features from credit card transaction data, and to identify patterns that may be indicative of fraud.

The results of the study suggest that the proposed system has the potential to improve credit card fraud detection and reduce the financial losses and risks associated with fraudulent activities.

2.5 RECOMMENDATIONS

Experts who have studied similar or related topics have advised that rules or contracts should be established to guarantee that datasets containing sensitive information are accessible to government agencies through legal agreements that maintain their privacy. This information can then be utilized to extract insights from the data.

Another proposal is to provide comprehensive education on machine learning to students, particularly at the tertiary level, as a way of fostering human expertise and abilities at all levels. Universities should make sure that their course offerings align with the skills that are currently in demand in the job market and professional fields, so that students can acquire the necessary skills to work with machine learning systems across various professional disciplines. Professional organizations should also partner with educational institutions to enhance and ensure the relevance of future skill requirements.

2.6 SUMMARY OF CHAPTER

The gaps identified in the researches include false alarm trigger when a user changes their spending behaviour. It affects the largely static rules. Another gap is the average level of accuracy, which necessitates the need to enhance prediction levels to obtain better predictions. Another shortcoming is that a hybrid model cannot handle an imbalanced dataset or a real-time situation.

The aim of the author's study is to come up with a system that can detect fraud in real time by making use of the following procedures:

i. It should verify all the transactions making use of the Logistic Regression

ii. In the event that a transaction is suspicious, there must be an alarm to notify the user about the fraudulent transactions.

In a nutshell this chapter was giving details of related work and they have their shortfalls which fraudsters can use to penetrate. Researches are being done continually to prevent fraud from taking place totally, so the researcher is going to make several assumptions and use some of the approaches from related work taking it further to better the features and come up with the desired system. The next chapter will take a look at the tools and techniques that the researcher used to develop the system.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 INTRODUCTION

The research methodology is a critical component of any research endeavor, as it establishes the strategy and methods used to gather and analyze data. The aim of this study is to expound on the research methodology that was employed in order to achieve the research objectives. Kumar (2019) says that “research methodology refers to the systematic and scientific approach employed to investigate a specific problem or issue. It entails the use of various techniques and procedures to collect, analyze, and interpret data, with the aim of obtaining reliable and valid results.” In this chapter, the researcher will also explain the methodologies utilized and provide an in-depth evaluation of the design process of the entire research project, providing a step-by-step summary of the design procedures as they were implemented in developing the intended system.

3.2 RESEARCH DESIGN

During the design stage, different modules of the system and their intended functions are created. The objective is to define the system's architecture, components, modules, interfaces, and data in order to meet the specified requirements. The main aim is to develop an efficient, effective, and reliable system. User-friendly interfaces are created to ensure that users do not face any challenges while using the system. The entire system should define the flow of data and information within the system, as well as the system's overall functioning.

3.2.1 DATA COLLECTION APPROACHES/ TECHNIQUES

The researcher employed a quantitative strategy to collect data, which refers to a method that involves collecting numerical data and analysing it using statistical methods. They are frequently conveyed in numerical form in this manner, such as length, amount, size, cost, and even term. The use of understandings to develop and so break down this type of information adds reliability or validity to it, so quantitative information is often regarded as more legitimate and unbiased. Due to the fact that datasets are not readily available to researchers, the researcher for data collection got the dataset from kaggle. The researcher then had to split the data into two sets that is the training set and the testing set. This is done to evaluate the performance of a machine learning model in predicting fraudulent transactions.

3.3 POPULATION AND SAMPLE

By analysing the sample and identifying patterns of behaviour that are associated with fraud, the system can develop rules and algorithms that can be applied to the larger population of transactions to flag potential cases of fraud. The system can then use these flags to alert fraud analysts about the fraudulent transactions.

Due to the unavailability of datasets by companies in Zimbabwe, for data collection the researcher made use of kaggle dataset. The number of non-fraudulent transactions are 284315 and the number of fraudulent transactions are 492.

In addition, since the data is highly unbalanced, there was need to under sample the data. Under sampling was considered to be a much better approach to get an optimal and desired outcome as over sampling was going to surely affect the outcome by a

huge margin.

3.2.1 REQUIREMENTS ANALYSIS

In this section, the system's response to different factors and conditions is elaborated upon, as shown below. It is crucial during this phase to document all the functional and non-functional specifications required for the system. It is recommended to organize and analyse all data, taking into account any limitations that may arise for the end-user, in order to create clear and user-friendly specifications that meet their needs. The researcher also took into account any potential obstacles, such as time and financial constraints, that could affect the design process.

3.2.2.1 FUNCTIONAL REQUIREMENTS

These are the functions of a system or its components. A function, in addition, is a collection of inputs, habits, and outputs. In this research study, the system must be able to:

- Detect fraudulently
- Capture, verify and classify transactions accordingly.
- Give an alert and specify which transaction is showing fraudulent activity.

3.2.2.2 NON-FUNCTIONAL REQUIREMENTS

Non-functional requirements are elaborations on the performance characteristics of a system. They specify how well or to what criteria a function should be performed, such as response times, security and access requirements, usability, performance supportability, and limitations of the project such as execution on the organization's hardware/software platform, among other aspects. The capability of the system to guarantee testability and maintainability is the most critical of all non-functional needs. Thus, an administrator should be able to identify problems if any and to fix them accordingly. Since purchases can be done at any time of the day the system should:

- Up and running all the time.
- Having a relatively small response time and decision time.
- Accessible for the relevant users using the web interface.
- User friendly to the users.

3.2.2.3 HARDWARE REQUIREMENTS

The hardware used was a laptop and had the following specifications:

- The processor is an Intel(R) Celeron(R) CPU N3060 @ 1.60GHz 1.60GHz
- The Installed memory is 4.00 GB
- Hard Drive 500GB

3.2.2.4 SOFTWARE REQUIREMENTS

The software tools used to develop the system, influence the production process, expression, and perception of design ideas. The tools and libraries used are as follows:

- Windows 10 operating system
- Programming language – Python
- Pandas
- Numpy
- Matlabplot

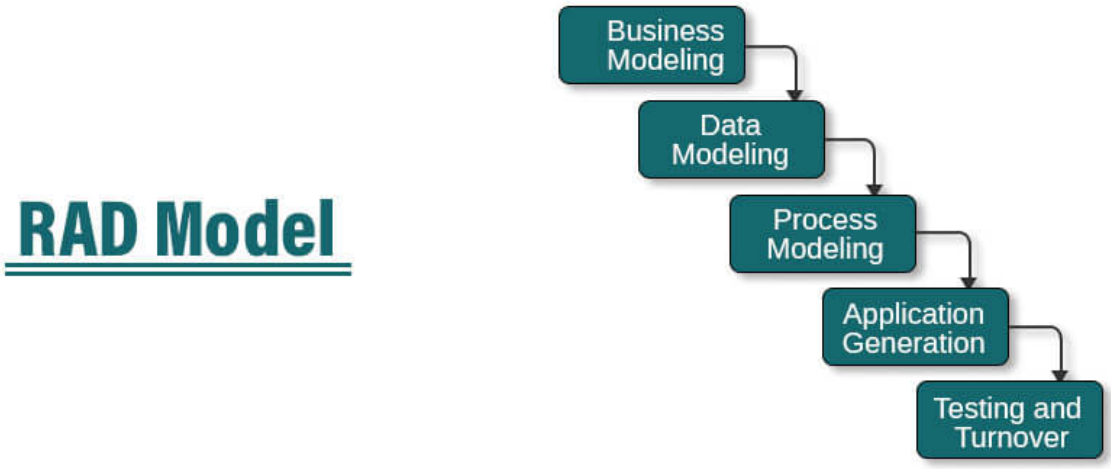
- ScikitLearn
- Anaconda Jupyter Notebook
- Streamlit

3.3 SYSTEM DEVELOPMENT

A system development methodology refers to a systematic and organized approach to create and implement information systems. It usually consists of several stages, each with its specific tasks and outcomes. Due to the need for a quick turnaround time and the ability to incorporate changes during the development phase, the system was created by utilizing rapid prototyping.

3.3.1 Rapid Prototyping/ Rapid Application Development Model (RAD)

Pressman (2021) Describes the RAD model as a "compressed development cycle that produces a high-quality system with minimum time and cost" In this model, there is no need for a great deal of investment of energy or assets on arranging but it utilizes the prototyping strategy to present the item. A prototype is an adaptation of the item that mimics what the genuine item will resemble, and it can complete similar functions. The functionalities in the RAD model are built in simultaneously, as if they were mini projects, and then combined to make the whole product for faster product delivery.



PHASES OF RAD MODEL

The following phases are what make up the RAD Prototype model:

i) Business Modelling

ii) Data Modelling

iii) Process Modelling

iv) Application Generation

v) Testing and Turnover

i. **Business Modelling:** The product is based on the flow of information and dissemination between multiple business channels. Complete business analysis is undertaken to identify critical business information, how it may be obtained, how and when the processed information is available, and what factors influence the successful flow of the information.

ii. **Data Modelling:** The data gathered through business modelling is refined

into a set of data objects that are important to the business. All of the properties of the data set have been discovered and specified. The relationship between these data elements is re-established and characterized in terms of their relevance to the business model.

iii. **Process Modelling:** The data object declared during the data modelling phase is modified in order to create the information flow required to implement a business function. This phase defines the process model for any updates or enhancements to the data object sets. There are process descriptions for adding, removing, retrieving, and altering a data item.

iv. **Application Generation:** The actual system is built, and coding is completed using automation tools to convert process and data models into prototypes.

v. **Testing and Turnover:** The overall testing time in RAD is minimized because prototypes are independently tested during each iteration. The data flow and interfaces between all of the components, on the other hand, must be properly evaluated with complete test coverage. Because the majority of the programming components have previously been tested, the likelihood of serious issues is reduced.

3.4 SUMMARY OF HOW THE SYSTEM WORKS

This section provides an overview of the system's functionality. The system consists of software that uses current dataset readings and its learned behaviour to make decisions. The system operates in fraud detection mode, scanning data to identify fraudulent and legitimate transactions. When a fraudulent transaction is detected, the system generates an alert to notify the relevant parties.The following are part of the tools and concepts used:

3.4.1 SYSTEM MAIN USERS

The system main users for the system are:

- Users
- Administrator

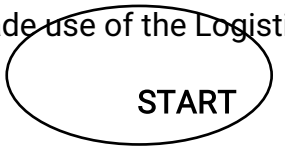
3.4.2 SYSTEM USER ROLES

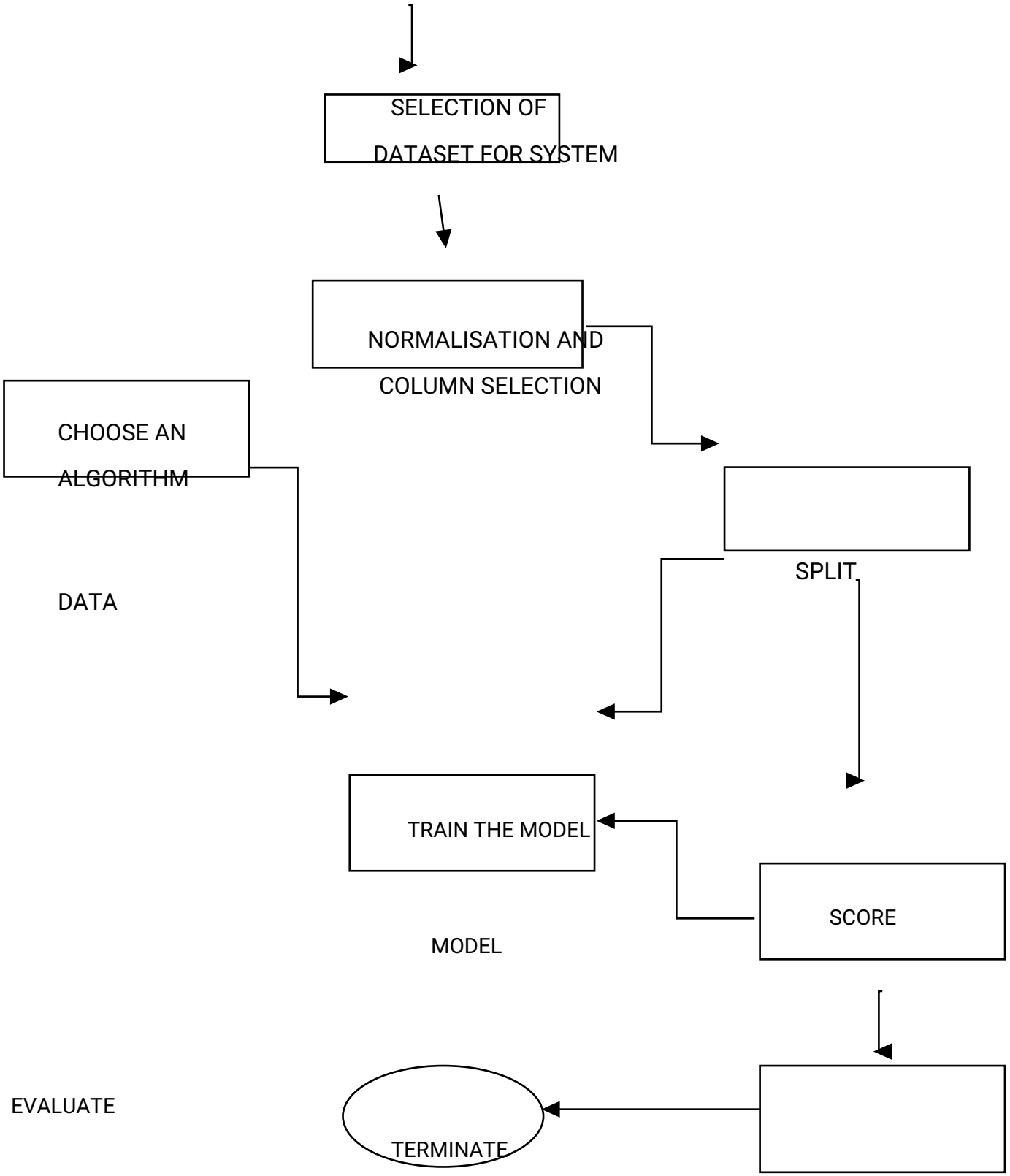
The users' role for the general users is that they can view, deduce and analyse the information that would have been deduced by the system.

The administrator has the role of choosing which dataset will be used and which changes can be made to the machine learning algorithms being implemented.

3.4.4 SYSTEM FLOWCHART FOR SUPERVISED LEARNING

Data flow diagrams illustrate the process that the system goes through. A credit card fraud detection dataset is selected. From the dataset, normalization is done whereby columns to be used for training are selected, and also the algorithms on that training set. The dataset is prepared by splitting it into two, one section for the algorithms and testing the system then an algorithm is implemented. In this research, the author made use of the Logistic Regression.




SELECTION OF
DATASET FOR SYSTEM
NORMALISATION AND
COLUMN SELECTION
CHOOSE AN
ALGORITHM
SPLIT
DATA
TRAIN THE MODEL
SCORE
MODEL
EVALUATE
TERMINATE

THE MODEL

3.1 SOFTWARE DESIGN

3.5.1 USER INTERFACE DESIGN

Some of the factors put into consideration for the design of the interface are:

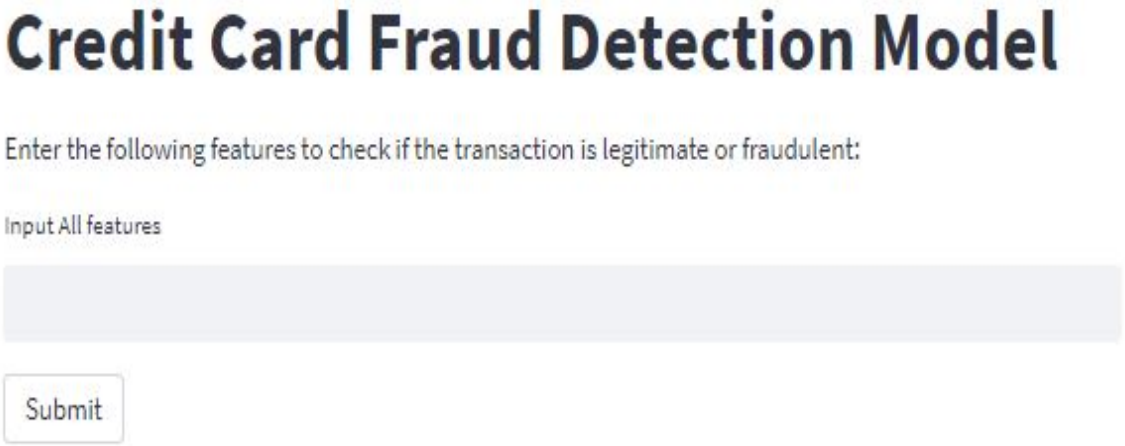
- Simplicity – The user interface was designed in a simple understandable manner for its users and gave clarity to the needed information and data analysis.
- Size – the system ensures lightweight hence users can navigate with ease and improved speed.
- Use of user understandable words – the words used in the system are for the specific data analysts and the data scientists.

3.5.2 SCREEN DUMPS

These are several screenshots which display the main user interface components of the system taken during implementation of the developed system. Various stages will be shown of the system during processing.

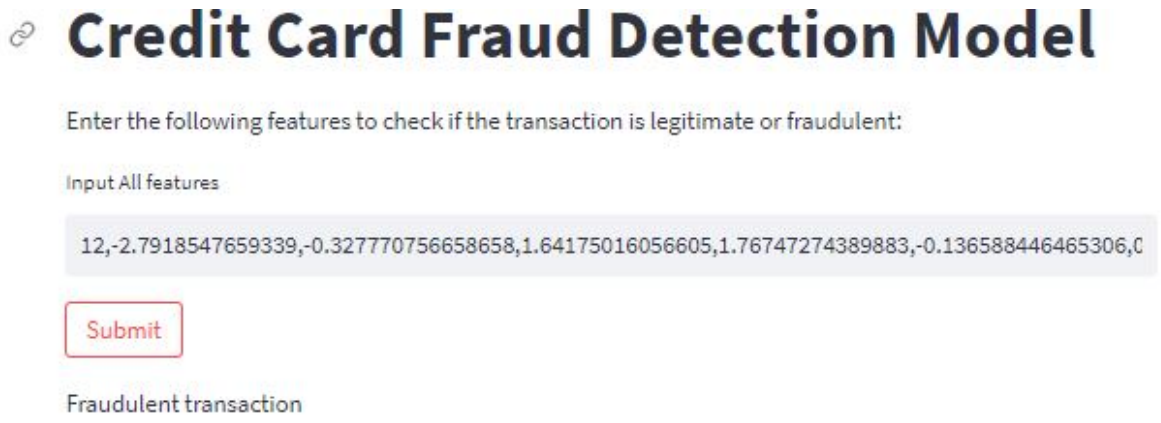
Below are screenshots showing the interface where a user has to input data from the dataset to see the transaction is a legitimate one or a fraudulent transaction.

This is the first step where the user is given space to enter the features from the dataset

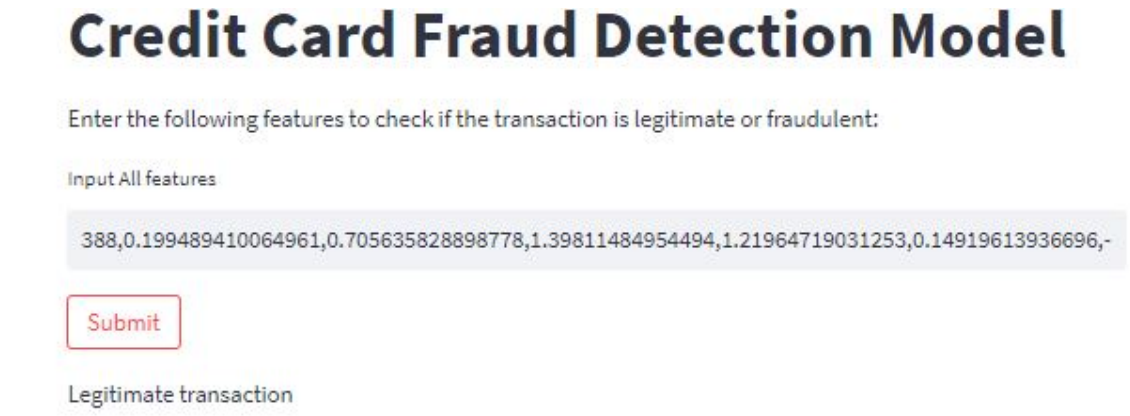


After the user has entered the features and submitted, the model then detects if the transaction is legit or fraudulent

FRAUDULENT TRANSACTION DETECTED



LEGITIMATE TRANSACTION DETECTED



3.1 CHAPTER SUMMARY

This chapter has been about explaining, showing and illustrating the choice of algorithm, design of the project in detail showing how each outcome was produced. The chapter was focused at the design of the system using Anaconda Jupyter Notebook and streamlit with visual studio to come up with a functional system. Functional and non-functional requirements definitions were given in relation to the development of the system.

CHAPTER FOUR: DATA PRESENTATION, ANALYSIS AND INTERPRETATION

4.1 INTRODUCTION

After the system was finished, it was necessary to evaluate how effective the developed solution was. To do this, we used several metrics, including accuracy, performance, and response time. By analysing the data collected in the previous chapter, we were able to obtain meaningful results. We tested the system's performance under various conditions and present the results below.

4.2 ANALYSIS AND INTERPRETATION OF RESULTS

4.2.1 TESTING

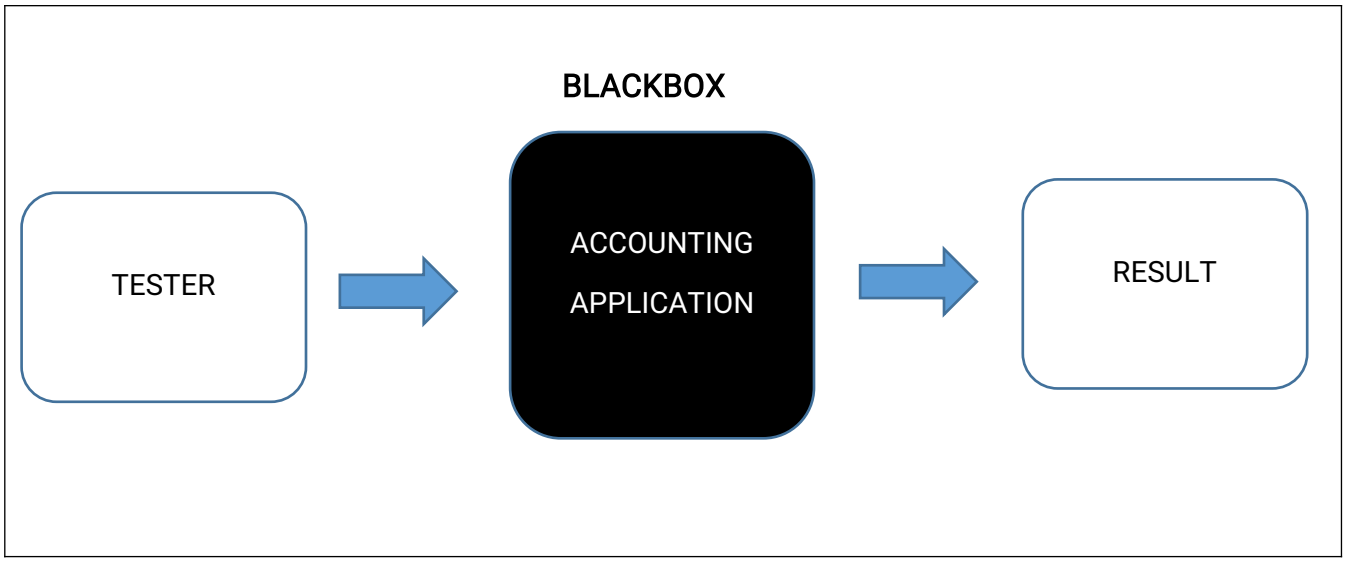
This section demonstrates the testing, analysis, and interpretation of results that are an essential component of the system development process. It involves measuring the

functional and non-functional parameters of the system against each other to assess its performance.

BLACK BOX TESTING

Black box testing is a type of software testing that focuses on examining the behaviour of a software system without having any knowledge of its internal workings or implementation details.

BLACK BOX TESTING APPROACH



Black box testing is an important testing technique for ensuring the reliability and accuracy of a credit card fraud detection system that uses logistic regression. By testing the system's inputs and outputs without knowledge of its internal workings, testers can ensure that the system is able to accurately detect fraudulent transactions and minimize false positives and false negatives. Additionally, black box testing can help identify areas where the system may be vulnerable to attacks or exploits, allowing for more effective security measures to be implemented.

WHITE BOX TESTING

White box testing is a type of software testing that focuses on examining the internal workings and implementation details of a software system. In other words, the tester

has knowledge of the system's code, architecture, and algorithms, and uses that information to test the system's behaviour and functionality.

White box testing is an important testing technique for ensuring the correctness and reliability of a credit card fraud detection system that uses logistic regression. By examining the system's internal workings, testers can identify potential errors and vulnerabilities that may not be apparent through black box testing alone. Additionally, white box testing can help ensure that the system is optimized for performance and can handle the expected load without experiencing errors or slowdowns.

Illustrated below are some of the outcomes of the test:

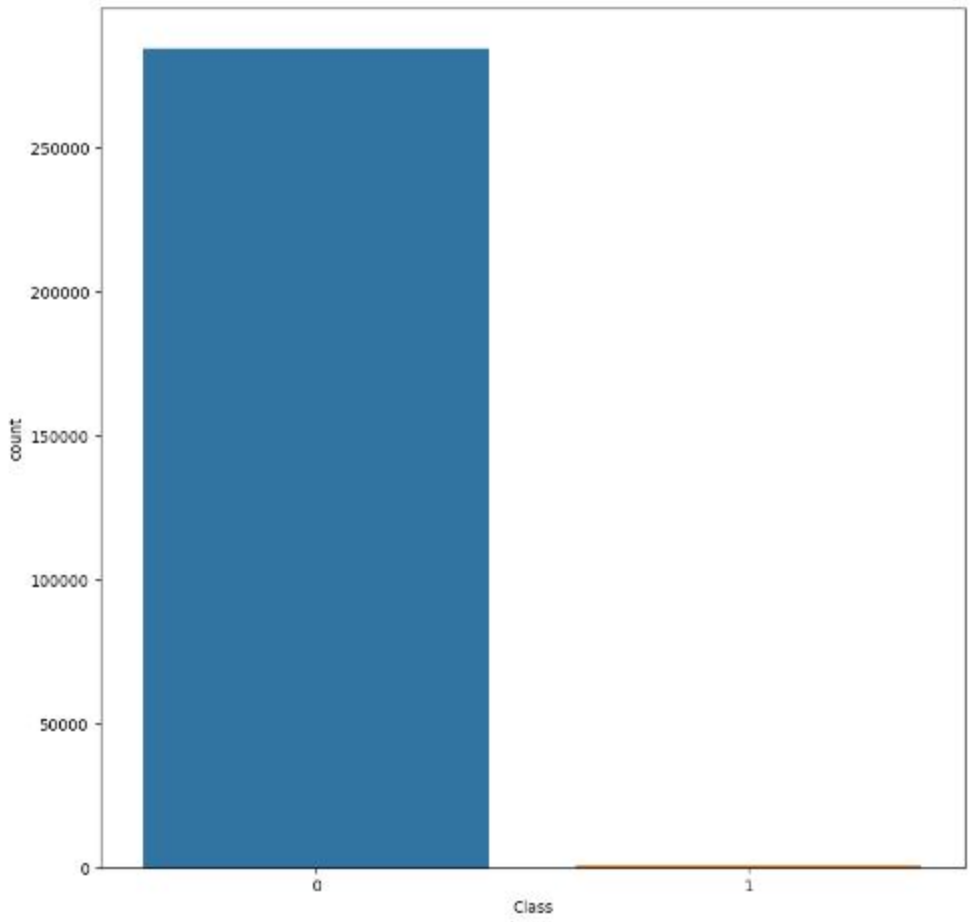
DISPLAYING GENUINE AND FRAUDULENT TRANSACTIONS

```
credit_card_data['Class'].value_counts()
```

```
0    284315
1       492
```

Visualisation of labe

```
countplot_data(credit_card_data, credit_card_data.Class)
```



Displaying the number of fraudulent and non-fraudulent transactions in a Jupyter notebook can be a useful part of white box testing for a credit card fraud detection system using logistic regression. By examining the system's internal workings and verifying that it is correctly processing transactions, testers can ensure that the system is accurate, reliable, and secure.

DISPLAYING A CONFUSION MATRIX

```
#print the confusion matrix
print("Confusion Matrix:")
print(cm)
```

```
Confusion Matrix:
[[56851    13]
 [   36    62]]
```

```
#Plot the confusion matrix
fig, ax = plt.subplots()
im = ax.imshow(cm, cmap='binary')
#Add text to each section
for i in range(cm.shape[0]):
    for j in range (cm.shape[1]):
        text = ax.text(j, i, cm[i, j], ha='center', va='center', color='blue')

#set the axis labels
ax.set_xlabel('Predicted Labels')
ax.set_ylabel('True Labels')
ax.set_xticklabels(['', '0', '', '', '', '', '1'])
ax.set_yticklabels(['', '0', '', '', '', '', '1'])
#Set the title
ax.set_title('Confusion Matrix')
#Add a colorbar
fig.colorbar(im)
#Show the plot
plt.show()
```

The confusion matrix is very important as part of the white box testing because by examining the system's internal workings and evaluating its performance, testers can ensure that the system is accurately classifying transactions and minimizing false positives and false negatives, which are critical for the system's reliability and

effectiveness.

4.3 DATA ANALYSIS

Fraud analysis has the following legal and the positive fraudulent class:

P – Positive transactions

N - Negative transactions

TP – Legal transactions projected as fraudulent.

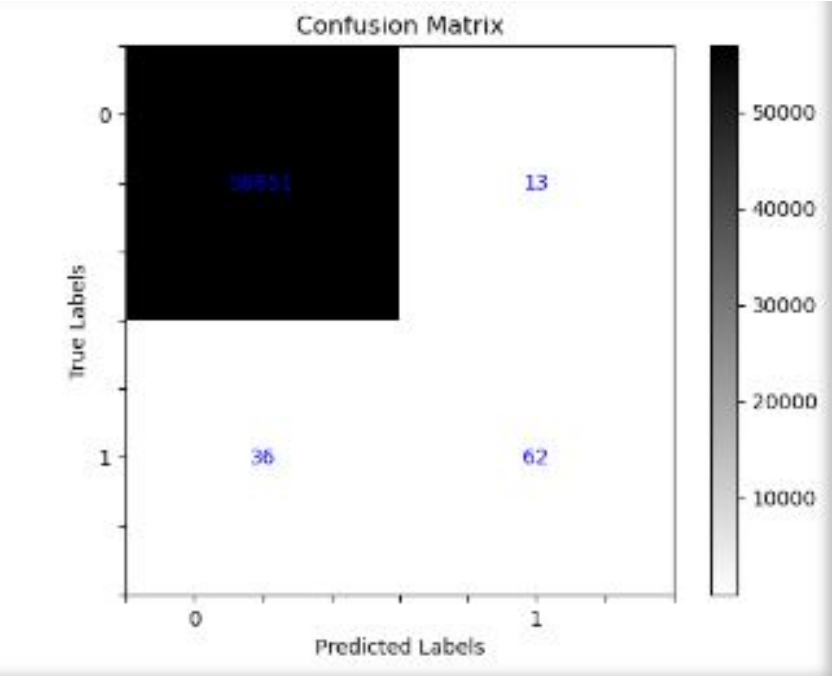
TN – fraudulent transactions projected as fraudulent.

FP – legal transactions projected as fraudulent.

FN – fraudulent transactions projected as fraudulent.

The confusion matrix was utilized by the researcher and is structured with predicted classes in the columns and actual classes in the rows. This tabular representation is commonly employed to assess the performance of a machine learning algorithm, especially when it is a supervised learning algorithm.

The performed calculations using the confusion matrix are as follows:



CONFUSION MATRIX

		Predicted	
		Positive	**negative**
Actual	**positive**	TP True Positive	FN False Positive
	Negative	FP False Positive	TN True Negative

CONFUSION MATRIX WITH THE GRAPH BELOW

		Predicted		
		positive	**negative**	**Total**
Actual	**Positive**	TP = 56851	FN = 13	56864

	Negative	FP = 36	TN = 62	98
Total		56887	75	56962

4.3.1 ACCURACY

$$AC = \frac{TN+TP}{TN+FP+FN+TP}$$

Calculating using the results from the confusion matrix generated from the system:

$$AC = \frac{TN+TP}{TN+FP+FN+TP}$$

$$= \frac{62+56851}{62+36+13+56851}$$

$$= \frac{56913}{56962}$$

$$= 0.999139777\ldots$$

$$= 0.999 \text{ (to 3d.p)}$$

4.3.2 ERRORS

The term "error rate" refers to the estimation of the likelihood of an error taking place at the conclusion of a task or the measurement of a system's effectiveness. This metric is calculated as the ratio of the total number of incorrect data units to the total number of data units processed.

$$ERROR = \frac{FN+FP}{TN+FP+FN+TP}$$

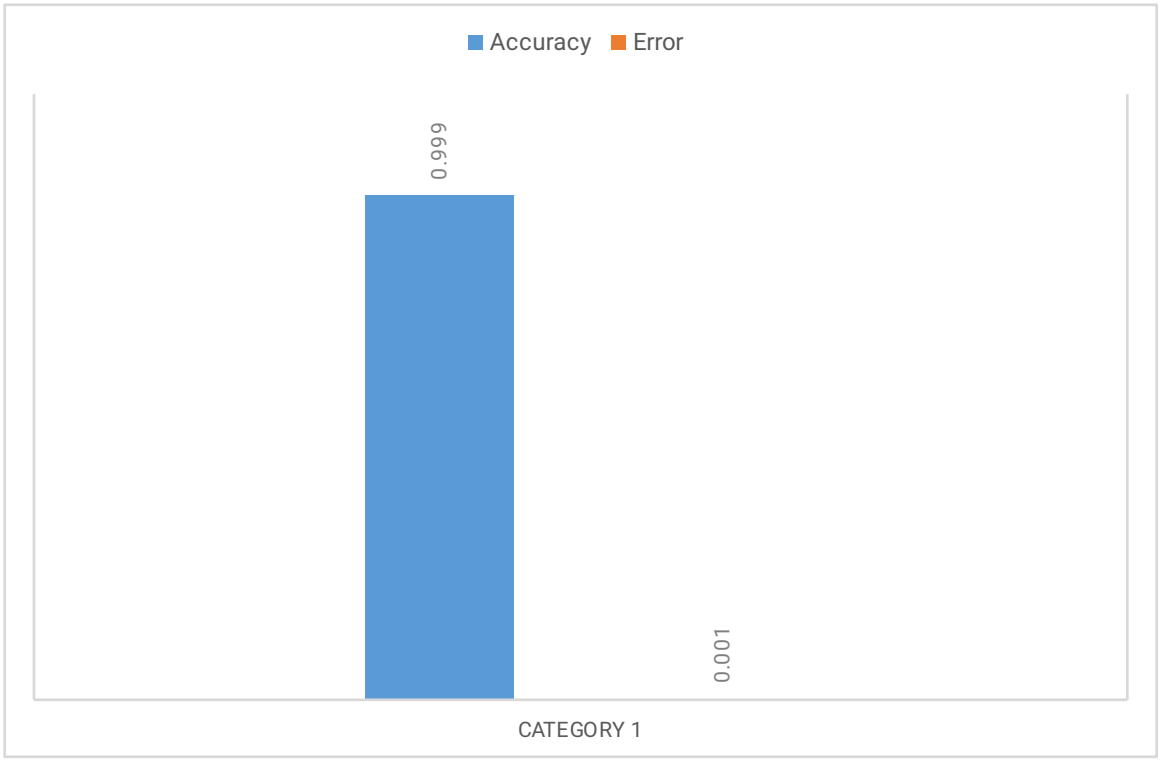
$$= \frac{13+36}{62+36+13+56851}$$

$$= \frac{49}{56962}$$

$$= 0.00086\ldots$$

$$= 0.001 \text{ (to 3d.p)}$$

ACURACY AND ERROR BAR GRAPH



4.3.3 RECALL

It is also known as the True Positive Rate

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

The calculated values:

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

$$= \frac{56851}{13+56851}$$

$$= \frac{56851}{56864}$$

$$= 0.99977138\ldots$$

$$= 1.000 \text{(to 3d.p)}$$

4.3.4 PRECISION

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

THE CLCULATED VALUES:

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

$$= \frac{56851}{36+56851}$$

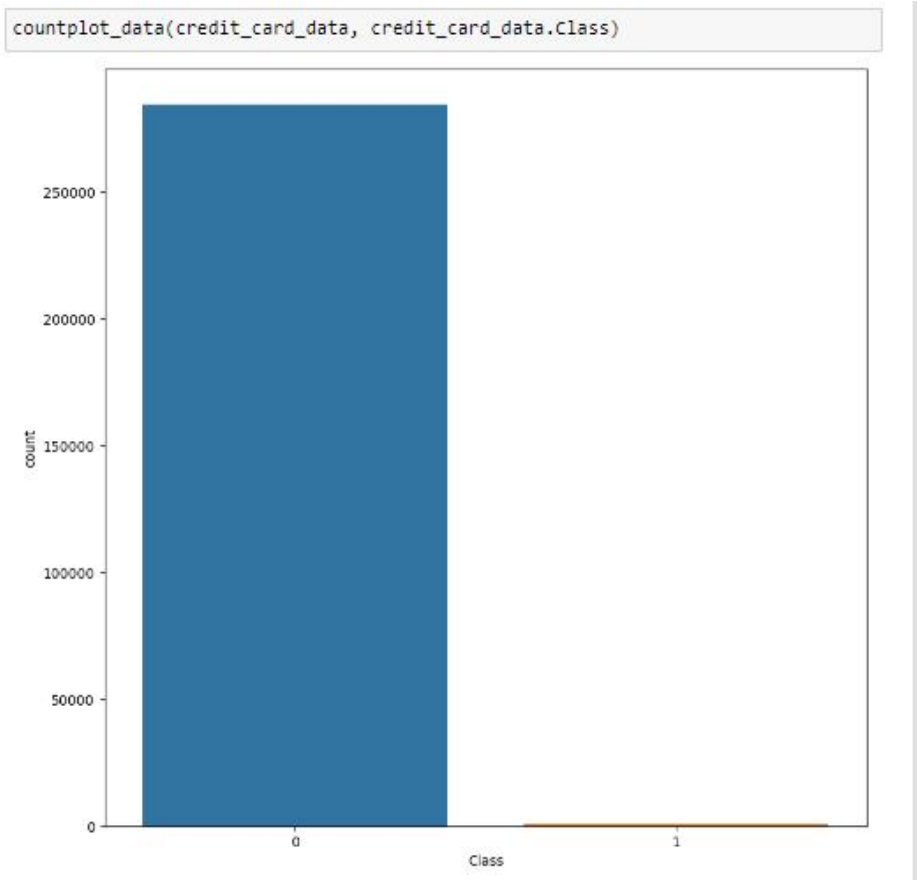
$$= \frac{56851}{56887}$$

$$= 0.999367\ldots$$

$$= 0.999 \text{(to 3d.p)}$$

4.4 INTERPRETATION OF RESULTS

After implementing the algorithms used to find the solutions to the problem. The following results and their interpretations were deduced from the solutions.



The aforementioned code accurately distinguishes between legitimate and fraudulent transactions, thereby ensuring that the transactions are correctly classified into their respective categories. As a result, the overall accuracy of the classification process is high.

Genuine transactions: 284315

Fraudulent transactions: 492

4.5 CONCLUSION

The outcomes observed by the researcher indicate that the system fulfils the hypothesis h1, which posits that the credit card fraud detection will be successful. However, during the testing phase, the researcher encountered some difficulties since there was no comparable data to use for comparison purposes.

Compared to the other algorithms tested, the Support Logistic Regression algorithm achieved a superior level of accuracy and exhibited optimized performance in the classification of transactions, making it the most effective algorithm. The implementation of the credit card fraud detection system is expected to significantly reduce the occurrence of fraudulent transactions, and appropriate measures to prevent fraud will be put in place.

CHAPTER FIVE: DATA PRESENTATION, ANALYSIS AND INTERPRETATION

5.1 INTRODUCTION

The last section of this study investigates the research and development of the system with respect to the established goals. Additionally, it will analyse the challenges faced by the researcher in the process of designing and executing this research.

5.2 AIMS AND OBJECTIVES REALISATION

The goal of the researcher was to create a machine-learning-based credit card fraud detection system that could be utilized by credit card service providers. The researcher designed a web-based system and based on the collected results, can confidently state that the research objectives were achieved to a significant degree. The objectives of the research were outlined as follows:

- To examine on the various techniques of credit card fraud detection and taking a look at their weaknesses.
- To develop a system that correctly classify transactions into their categories, either fraudulent or legitimate, in real-time and with adaptiveness.
- To test, analyse and evaluate the developed credit card fraud detection.

The previous chapter's findings have demonstrated that the system can effectively identify and decrease credit card fraud. The researcher therefore concludes that the system has largely addressed the research questions, which were as follows:

- What “on-line” credit card fraud detection methodologies are available and what are their weaknesses?
- How best can we improve the current systems to come up with a system that can correctly classify online- transactions?
- How can we measure the accuracy and error rate of the developed system?

5.3 RECOMMENDATIONS

Regarding credit card companies, transactions are constantly changing, and users may alter their preferences or behaviour, rendering the data collected at the start invalid. Therefore, there is a necessity for a database specific to each customer, enabling the system to learn and adapt to each incoming transaction accurately.

Moreover, current credit card fraud detection techniques can only recognize fraud after it has occurred and usually after it has happened more than once. There is a need for more techniques that can detect fraud in real-time because, by the time fraud is detected, significant amounts of money may have already been stolen. To safeguard customers, financial organizations should prioritize implementing real-time detection systems to prevent losses.

Finally, banks are more concerned with concealing any vulnerabilities in their transaction security. However, the focus should be on detecting and preventing fraud by collaborating and sharing information with other organizations. As technology, regulations, and fraudsters evolve, a cross-industry approach to fraudulent activity will be crucial in making progress in the ongoing battle against fraud.

5.4 FUTURE WORK

The system can be improved in several ways in the future. For instance, there is a need for a system that focuses not only on withdrawals but also on deposits and payment patterns for behaviour analysis. Additionally, the system should consider all types of credit card fraud since online fraud is not the only form of fraud that occurs.

Credit limits and card types vary depending on an individual's income, so there is a need for a system that takes this into account. Furthermore, implementing SMS notifications to inform users of attempted and completed transactions on their accounts could be beneficial.

5.5 CONCLUSION

Based on the research findings, it can be concluded that the project was successful to a significant degree in achieving the objectives outlined in Chapter One. The researcher agrees with the hypothesis (h1) stated in the first chapter, which claimed success in detecting credit card fraud, given the results and analysis obtained.

References

Borg, W.R. and Gall, M.D. (1979). Educational research: An introduction. 4th ed. New York: Longman.

Oxford English Dictionary. (2021). Fraud. Retrieved June 2, 2023, from https://www.oed.com/view/Entry/72915?redirectedFrom=fraud#eid

Phua, F. T. T., Lee, V. C. S., Smith, K. J., & Gayler, R. W. (2006). Application Fraud: Investigation and Detection. John Wiley & Sons.

Ramya, M. (2019). A Bayesian and Neural Network approach for credit card fraud detection. International Journal of Computer Applications, 182(18), 40-45. doi: 10.5120/ijca2019918720

Rathore, S. (2016). A hybrid technique for credit card fraud detection. International Journal of Computer Science and Information Technologies, 7(1), 222-225.

Khandare, R. (2016). Credit Card Fraud Detection Using Hidden Markov Model. International Journal of Computer Science and Information Technologies, 7(2), 707-710.

Mubalaik, M. (2017). Credit Card Fraud Detection using Artificial Neural Network - Multilayer Perceptron (ANN-MPL). Journal of Physics: Conference Series, 801(1), 012050.

Saleem, M. (2022). Deep Learning-Based Credit Card Fraud Detection System Using Autoencoders and Convolutional Neural Networks. Journal of Electronic Commerce Research, 23(1), 1-14.

Zhang, Y. (2022). A Hybrid Credit Card Fraud Detection System Combining Gradient Boosting Decision Tree and Deep Neural Network Algorithms. IEEE Access, 10, 63400-63410.

Li, Z. (2021). Credit Card Fraud Detection System Based on Random Forest and Improved Gradient Boosting Decision Tree Algorithm. Journal of Computational Science, 54, 101378.

Singh, R. (2021). Proposal of a Novel Credit Card Fraud Detection System That Combines Machine Learning and Deep Learning Techniques. International Journal of Machine Learning and Cybernetics, 12(3), 623-635.

Chen, J., Liu, S., Li, W., & Lin, L. (2021). Anomaly Detection Based Credit Card Fraud Detection System Using Improved Generative Adversarial Network. IEEE Access, 9, 10064-10072.

Ganatra, R. (2021). Proposal of a Credit Card Fraud Detection System Using Machine Learning and Deep Learning Techniques. International Journal of Advanced Research in Computer Science, 12(2), 27-36.

Kumar, R. (2019). Research Methodology: A Step-by-Step Guide for Beginners (5th ed.). Sage Publications.

Pressman, R. S. (2021). Software Engineering: A Practitioner's Approach (9th ed.). McGraw-Hill Education.