

**BINDURA UNIVERSITY OF SCIENCE EDUCATION  
FACULTY OF SCIENCE  
COMPUTER SCIENCE DEPARTMENT**



**CREDIT CARD FRAUD DETECTION SYSTEM**

**BY**

**BERYL ADAMS**

**B1335452**

**SUPERVISOR: MR CHIKWIRIRO**

*A thesis submitted in partial fulfilment of the requirements for the Bachelor of Science  
Honours Degree in Computer Science.*

**JULY 2017**

## Approval Form

The undersigned certify that they supervised the student Beryl Adams (B1335452) dissertation entitled “**Credit Card Fraud Detection**” submitted in Partial fulfilment of the requirements for the Bachelor of Computer Science Honours Degree of Bindura University of Science Education.

.....

**STUDENT NAME**

.....

**DATE**

.....

**SUPERVISOR**

.....

**DATE**

.....

**CHAIRPERSON**

.....

**DATE**

.....

**EXTERNAL EXAMINER**

.....

**DATE**

## **Declaration**

I declare that the knowledge, attitudes and perceptions regarding the credit card fraud detection system is my own particular work and that all sources that I have utilized or consulted have been revealed and acknowledged by methods of complete reference.

“The cure for boredom is curiosity. There is no cure for curiosity.”

Dorothy Parker

## **Dedication**

This dissertation is dedicated with love to my father Sondon Adams and Mercy Adams my mother for being there since the beginning of time with their prayers, love, as well as support. I would also like to dedicate this dissertation to my siblings as well, Robin, Andy, Heather and Chloe for being there always, I am forever grateful for their love. I appreciate the sacrifices they made towards my personal and professional endeavors. May God dearly bless them all.

## **Acknowledgement**

First and foremost, I would like to thank the Lord Almighty for everything including His unending grace, favor and mercy for because of Him I have reached this far. Further acknowledgements go to my brothers, sisters and dear parents for the unconditional love and support throughout the journey.

I also wish to express my heartfelt gratitude to those whose support made the completion of this study possible. In particular I would like to thank MR. CHIKWIRIRO for his support, patience and wisdom may God bless him. His help urged me to work unbiasedly and usefully on this research and towards the achievements of my objectives.

In addition, I want to acknowledge the constructive criticism from my colleagues, Kushinga Mukabeta, Aleck Chipato, Perseverance Mudzinganyama, Terence Murowanedzo, Samuel Masuka as well as Kudzaishe Machivenyika just to mention a few.

To sum up, special gratitude goes to Bindura University of Science Education and all the Computer Science Department lecturers for nurturing me and giving me the opportunity to come up with this project in fulfillment of the requirements of the Bachelor of Science Computer Science honors degree.

## **Abstract**

A lot of money is lost every day due to fraudsters so it is vital to come up with algorithms that can be keys to reduction of these losses. The design of fraud detection algorithms is however particularly challenging due to the non-stationary distribution of the data, the highly unbalanced classes distributions and the availability of few transactions labelled by fraud investigators. At the same time, public data is scarcely available due to confidentiality issues, leaving unanswered questions about the best strategy. Detection Systems should be able to meet real-world working conditions and also be able to integrate researchers' feedback to generate accurate alerts. Credit cards are becoming the most popular forms of payment for online and offline. Due to the raise in the use of credit cards worldwide fraudsters are also increasing.

Credit cards provide cashless shopping at every shop in the world. In Credit card fraud detection system, fraudulent transaction will be detected after a transaction is done. Credit card fraud can be detected using the Support Vector Machine algorithm, it is a Binary classification technique that can be used during "on-line" transacting. It is the tool for solving hidden problems. In this paper, the sequence of operations in credit card transaction processing system using Support Vector Machine and how it is used for fraud detection is shown. Using Support Vector Machine, the fraud detection processing system is trained with the standard procedures and spending patterns of a card user.

## Table of Contents

Approval Form.....	i
Declaration.....	ii
Dedication.....	iv
Acknowledgement .....	v
Abstract .....	vi
List of Figures .....	x
List of Tables .....	x
Chapter One: Problem Identification .....	1
1.0 Introduction.....	1
1.1 Background of the Problem.....	2
1.2 Problem statement.....	3
1.3 Objectives.....	4
1.4 Research Questions .....	4
1.5 Research Propositions.....	4
1.6 Justification of study .....	4
1.7 Assumptions .....	5
1.8 Limitations and challenges .....	5
1.9 Scope of research/Delimitations .....	5
1.10 Chapter Summary .....	5
Chapter Two: Literature Review .....	6
2.0 Introduction.....	6
2.1 Relevant theory on the subject matter.....	6
2.2 Characteristics of Credit Card Fraud Detection systems.....	9
2.3 Data Mining and Detection Techniques.....	10
2.3.1 Data mining.....	10
2.3.1 Detection Techniques.....	10
2.3.2 Unsupervised Techniques.....	11
2.3.3 Supervised techniques .....	11
2.4 Existing research about Credit Card Fraud Detection systems .....	14
2.5 Inconsistencies and other shortcomings in our knowledge and understanding.....	18
2.6 Reasons for studying the problem further.....	18



2.7 Contribution that the present study expected to make .....	19
2.8 Chapter Summary .....	20
Chapter Three: Methodology .....	21
3.0 Introduction.....	21
3.1 Research Design.....	21
3.2 Data Collection Method .....	22
3.2.0 Quantitative Approach .....	22
3.3 System Development Methodology .....	22
3.4 Requirements Specifications .....	23
3.4.0 Functional requirements .....	23
3.4.1 Non-functional requirements .....	24
3.5 Design Tools .....	25
3.5.0 Hardware design tools .....	25
3.5.1 Software design tools.....	25
3.6 Description of the System .....	26
3.6.0 System main users .....	26
3.6.1 Use Case Diagrams.....	27
3.6.2 System flow chart for Supervised Learning. ....	28
3.6.3 System flow chart for Credit Card Fraud detection using Supervised Learning .....	29
3.6.4 The Relational Database Schema.....	30
3.7 Software Design.....	31
3.7.0 User Interface Design .....	31
3.7.1 Screen Dumps .....	32
3.8 Chapter Summary .....	35
Chapter Four: Data Analysis Presentation, Analysis and Interpretation .....	36
4.0 Introduction.....	36
4.1 Data Analysis .....	36
4.1.0 Accuracy Rate .....	38
4.1.1 Error Rate .....	38
4.1.2 Recall Rate .....	39
4.1.3 Precision Rate .....	39
4.2 Software Testing.....	40
4.2.0 Black box Testing.....	40

4.2.1 Fuzzy Testing.....	40
4.2.2 Alpha Testing .....	40
4.2.3 Beta Testing .....	41
4.2.4 Performance Testing .....	41
4.3 Chapter Conclusion.....	41
Chapter Five: Conclusions and Recommendations .....	42
5.0 Introduction.....	42
5.1 Aims and Objective Realization .....	42
5.2 Challenges Encountered .....	43
5.3 Recommendations.....	43
5.4 Future Work .....	44
5.5 Conclusion .....	44
References .....	45

## List of Figures

Figure 1: Outlier Diagram.....	9
Figure 2: RAD Diagram.....	23
Figure 3: System Use Cases.....	27
Figure 4: Supervised learning flow chart.....	28
Figure 5: Overall system flow chart.....	29
Figure 6: ER diagram.....	30
Figure 7: Account creation and login in platform.....	32
Figure 8: Account history. ....	33
Figure 9: Online Shopping mall diagram.....	34
Figure 10: Administrator interface.....	35
Figure 11: Confusion Matrix Graph .....	37
Figure 12: Accuracy and Error rate diagram. ....	38

## List of Tables

Table 1: Confusion Matrix.....	37
--------------------------------	----

## **Chapter One: Problem Identification**

### **1.0 Introduction**

For some time, there has been a strong interest in the ethics of banking (Molyneaux, 2007; George, 1992), as well as the moral complexity of fraudulent behavior (Clark, 1994). Fraud refers to obtaining goods or services and money using illegal ways. It can occur with any type of credit products, such as personal loans, home loans, extravagant spending, and retail among other things. Credit card fraud is defined as the illegal use of any system or, criminal activity through the use of physical cards or card information without the knowledge of the cardholder.

A Credit card is a standard-issue plastic token, with a magnetic stripe that holds a machine clear code. It is an advantageous substitute for money or check, and a fundamental segment of electronic business and web trade. Card holders draw on a credit restraint endorsed by the card-guarantor, for example, a bank, store, or specialist organization an aircraft, for instance. Cardholders typically should pay for their purchases inside 30 days of procurement to keep away from interest as well as penalties. The service provider generally a bank makes a spinning record and allows a credit extension to the cardholder, from which the cardholder can obtain cash for instalments to a trader or as a loan.

Quite a number of systems or models, have processes and preventive measures that try to help in stopping credit card fraud and reduce financial risks. Banks and credit card companies gather large amounts of credit card account transactions. Fraud detection is then based on analysing these credit card transactions for any unusual behaviour so as to reduce the rate of credit card frauds. However, with the developments in the Information Technology and improvements in the communication channels, fraud is spreading all over the world with results of large amounts of fraudulent loss.

Anderson (2007) identified and described the different types of fraud. Credit card frauds can be viewed in many different ways such as simple theft, counterfeit cards, Never Received Issue (NRI), application fraud and online or Electronic fraud where the card holder is not present. The

researcher will take a look at various techniques and come up with a system that will detect fraud online.

## **1.1 Background of the Problem**

Dating back to the mid 1990's, it was the beginning of real commerce from the Internet. With the start of e-commerce back in 1994 the first true buy buttons appeared on the Internet. Soon after several types of fraud emerged. The first fraud trend was the use of well-known names for example celebrities to commit fraud. Fraudster would use third-party stolen credit cards with the name of the celebrity of the day. Only completion of an authorization was required, the name used in the purchase was not checked and fraudsters used this to their advantage. Merchants got smarter and implemented rules to check the name being used. Since there are so many possible names, and so many people with the same name, it was only a temporary solution. Likewise, the fraudsters moved on to new attacks.

Next the fraudsters came up with card-generator applications for real credit card numbers, using these they defrauded merchants over and over again. As time progressed new forms of fraud emerged each time solutions were brought forward.

After 1996 fraudsters started to use the Internet as a test bed for stolen credit cards. The fraudsters could now test the credit cards and go on a shopping rampage over the internet up to date. In addition, a 2015 research note from Barclays stated that the U.S. is responsible for 47 percent of the world's card fraud in spite of only accounting for 24 percent of total worldwide card volume. The high level of debit and credit card fraud in the United States impacts other countries. Across the border fraud occurs when fraudsters use a consumer's credit or debit card data in one country to make fraudulent transactions in another country. That may finally be what it takes for us to guard ourselves, as the number of fraud victims is estimated to reach 14 million in 2018. The number of clients who experience a credit card breach and fraud in the same year is expected to raise 34 percent between 2014 and 2018.

## **1.2 Problem statement**

Organizations are constantly developing and investing in tools, resources and techniques for detecting and preventing fraud. One of the most common forms for prevention and detection mechanism is behaviour-based transaction monitoring. This involves analysing transactions, within the context of a customer's regular behaviour, to separate and alert the authorities in case of suspicious activity. Most of the solutions for detecting depend on the use of 'if-then' statements and the transactions will fall in an inflexible set of procedures. Since these systems have no flexibility they only conform to pre-programmed parameters hence fraudsters manipulate this to work around the system.

Even though most merchants have real-time risk scoring, this may not be enough since financial criminals can just attack from a low-profile risk score as well as a high-risk. Therefore, there is a need for real-time risk scoring that uses behaviour-based observing as a continuous feedback loop so as to validate profile risk scores.

Moreover, detection of fraud is a complex computational task and most systems just predict the likelihood of fraud, a good fraud detection system should identify the frauds accurately, it should detect fraud quickly and it should also not classify a genuine transaction as fraud or a fraudulent one as genuine.

### **1.3 Objectives**

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

- To examine on the various techniques of “on-line” 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, analyze 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 Propositions**

The research propositions are:

- $H_0$ : There will be no success in the detection of fraud.
- $H_1$ : There will be success in the detection of fraud.

### **1.6 Justification of study**

Since the introduction of credit cards, banks, credit card issuing companies as well as its people have been incurring unnecessary expenses due to fraud on their credit cards. This research

project will be aiming at coming up with a system that can ensure security and user satisfaction by identifying fraud in real time and preventing the transactions from being authorized in the first place as most techniques only detect fraud after a number of transactions have already been made.

### **1.7 Assumptions**

This study was based on the following assumptions

- Accessing the necessary information including datasets is not going to be a problem.
- Datasets can be universal, they are not constrained to its particular country of origin and it can be used anywhere.
- That the limit for the credit cards is the same for all card holders

### **1.8 Limitations and challenges**

- There were challenges to datasets access.
- Implementation of the system on a real platform was not possible.
- There was a limited amount of time to gather information and observe trends.

### **1.9 Scope of research/Delimitations**

The research project is intended for banks and service providers. New users would fill in the necessary details, old ones would make the necessary purchases and have their transactions monitored. Administrators at the organizations can view fraud, and make necessary changes if any are needed.

### **1.10 Chapter Summary**

The author in this chapter introduced the whole project and identified the problem. In the next chapter, the author will give a detailed literature review of what took place and what has been done so far regarding the research topic.



## **Chapter Two: Literature Review**

### **2.0 Introduction**

The review of related literature helps the researcher in integrating the current study to research studies on the specific topics that are similar to his or her study. It also attempts to identify what research has been done, being done and what needs to be done in the research study. (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." In relation to the above thesis (Borg & Gall, 2010) concurred with (Leedy, et al., 2009) that one should try to read the latest materials because continuous research is continuously bringing out new information.

Fraud is one of the major moral issues in the credit card industry. The primary goal for this section is, to survey elective methods that have been utilized as part of credit card fraud detection. The sub-point is to present, look at and the analyze and compare as of late distributed discoveries in the detection of credit card fraud, in the process identify inconsistencies as well as overall shortcomings and the contributions the current study is supposed to make.

### **2.1 Relevant theory on the subject matter**

The Oxford English Dictionary characterizes fraud as wrongful or criminal trickiness with the aim to bring about budgetary or individual gain. It is a growing problem all over the world nowadays. Fraud has changed significantly during the most recent couple of decades as advancements have been changed and created (Sharma, 2012). A basic assignment to help organizations, and monetary foundations including banks is to find a way to predict fraud and to manage it productively and viably, when it happens (Anderson, 2007).

Credit Card Fraud is defined as when someone uses another person's credit card for individual reasons while the proprietor of the card and the card issuer are not aware of the fact that the card is being used, the persons using the card has no connections at all with the proprietor of the card

or the issuer and has no plans of repaying for the costs incurred (Journal & Trends, 2014).

Fraud detection is where by fraud is identified as fast as possible as soon as it has taken place. Methods of fraud detection are continuously developed to prevent criminals from adapting to the detection strategies. Unfortunately, the process of developing new fraud detection methods is a complex task since there are extreme impediments in the trading of thoughts in the credit card fraud detection field. Datasets are not available and results are regularly not discharged to general society. The fraud cases must be recognized from the accessible immense datasets, for example, the logged information and client behavior. At the moment, fraud detection has been actualized by various strategies, for example, artificial intelligence, data mining and statistics. Fraud is found from peculiarities in information and patterns.

Transactions which are genuine are usually mixed up with transactions which are not genuine at all and it is very difficult to detect fraud accurately if simple pattern matching techniques are used. There are various technologies of credit card fraud detection which are employed by credit card issues companies and banks in order to minimize the loses. Some of the technologies are based on Artificial Intelligence, Data mining, Neural Network, Bayesian Network, Fuzzy logic, Artificial Immune System, K- nearest neighbor algorithm, Support Vector Machine, Decision Tree, Fuzzy Logic Based System, Machine learning, Sequence Alignment, Genetic Programming etc. This chapter will present a survey of various techniques used in credit card fraud detection mechanisms.

There is a lot of literature on a wide range of security methods to look at transactions from unauthorized users or exposure of private information and consequent valuable resources are available. Regardless of the measures which are always taken fraudsters always find a way to crack through the various credit card fraud detection techniques. Other forms of purchasing ways such as the use of debit cards and swiping on the ATM require that the user input a pin or a password for authentication in order to avoid fraud. However, credit cards work in a somewhat different manner in that it does not require a pin from the user but just the name, expiration date and account number of the user.

Credit card fraud can be divided into two types thus Offline fraud and On-line fraud, the former is done when a fraudster steals a physical card, the latter takes place via the web or internet (“Review Paper on Credit Card Fraud Detection,” 2013).

There are various forms of fraud, there is Theft Fraud/ Counterfeit Fraud, focus is on theft and counterfeit fraud, which are related to one other. Theft fraud refers to an action where someone uses a card that is not theirs. In the event that such a crime happens the owner should report to the bank or the credit card issuing company and measures will be taken. Whereas in counterfeit the credit card is used remotely since only credit card information is required.

Telecommunication Fraud is whereby a fraudster uses telecommunication services to commit other forms of fraud being discussed and usually businesses, communication services and businesses are affected.

There is also Computer Intrusion which is defined as the act of entering without warrant or invitation, that means potential possibility of unauthorized attempt to access information, manipulation of information purposefully. Intruders may be from any environment, an outsider a Hacker that is and an insider who knows the layout of the system.

Bankruptcy Fraud is whereby a credit card is used while owner is being absent and it is one of the most complicated types of fraud to predict or detect.

Application Fraud is where by a person makes an application to the bank or credit card issuing company for a credit card making use of false details. For detecting application fraud, two different situations have to be classified. When applications come from the same user with the same details, that is called duplicates, and when applications come from different individuals with similar details, it is called identity fraud. Phua et al. (2006) describes application fraud as the demonstration of identity crime, it occurs when application forms contain possible, and synthetic (identity fraud), or real but also stolen identity information (identity theft).

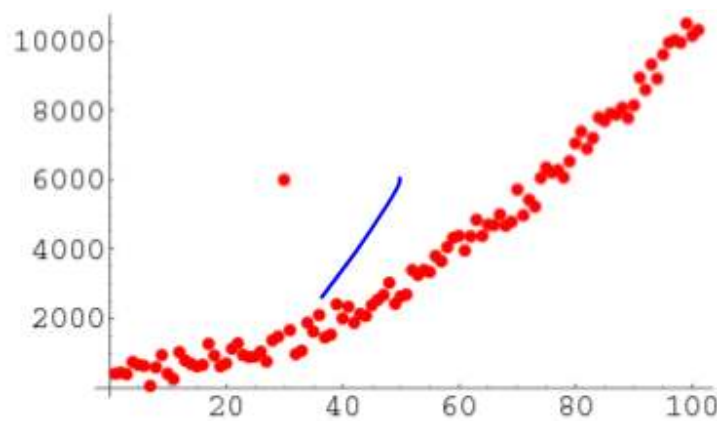
The main focus of the project is however on “on-line” credit card fraud detection.

## 2.2 Characteristics of Credit Card Fraud Detection systems

The detection of fraud is a complex task and there is no system that correctly predicts any transaction as fraudulent. The properties for a good credit card fraud detection system are that it should be able to:

- identify the frauds accurately.
- detect the frauds quickly.
- separate and classify a genuine transaction from fraudulent ones (Jayant, 2014).

Hawkins et al. (2002) define an outlier as “an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism”. Outlier detection is a critical task as outliers indicate abnormal running conditions from which significant performance degradation may happen. As shown below on the diagram the lone red dot is an outlier since it is torn away from the other red dots.



*Figure 1: Outlier Diagram.*

(Sciences, 2013)

## **2.3 Data Mining and Detection Techniques**

### **2.3.1 Data mining**

Data mining refers to a family of machine learning techniques which are capable of analyzing and extracting non-trivial patterns in given data. Since this technique can reveal previously unknown information hidden in data of various databases it is also referred to as knowledge discover. Hormozi et al. (2004) postulated that “data mining enables an organization to focus on the most important information in the database, which allows managers to make more knowledge decisions by predicting future trends and behaviors”. Since databases to organizations are extremely large it is time consuming and nearly impossible to do check manually for hidden patterns on data. Therefore, data mining can be introduced to facilitate the discovery of useful knowledge.

Srivastava et al. (2008) stated that the only way to detect credit card fraud is by analyzing the spending behavior of customers using data mining techniques. Customers tend to follow a standard spending profile and therefore any transaction which deviates from that standard can be considered as suspicious. Suspicious transactions can be examined in detailed by bank officers to determine whether they are indeed fraudulent or not. Like most of the machine learning algorithms, data mining techniques tend to learn models from data.

### **2.3.1 Detection Techniques**

Techniques used in fraud detection are divided into supervised and unsupervised techniques. The former uses past fraud cases to build models which will help in producing a suspicion score for new transactions. Moreover, supervised methods are only trained to differentiate between legitimate transactions and previously known fraud, where as in unsupervised techniques there are no previous data sets in which transactions can be known as fraudulent or legitimate. However, the later does not make any assumptions about the availability of labeled data, it's useful in applications where there is no prior knowledge about the particular class of observations in a data set. It simply seeks those accounts, customer etc., whose behavior is “unusual”. Some of the methods under each technique are given below (Jayant, 2014).

### 2.3.2 Unsupervised Techniques

**Peer Group Analysis:** this is the method used for monitoring an account's behavior over a period of time in data mining. If there is a change in the behavior of individual objects from other objects that they were behaving similarly to, the PGA notices that. An object is selected to be the target, it is compared to other objects that are in the database by summarizing previous behavior patterns of each object. A peer group of objects most similar to the target object is chosen by making use of comparisons. PGA is a part of the data mining process which involves cycling between the detection of objects that behave in suspicious ways and the detailed examination of those objects.

**Break Point Analysis:** this is a tool used in the detection of behavioral fraud. It is an observation or time given for detecting anomalous behavior in transactions. If there is a change in the sequence of transactions being compared for a particular account it is detected. Break point analysis makes use of a fixed length moving window of transactions, when a transaction is made it enters the window and old transactions are removed from the window. Its main advantage noted is that balanced data is not required as the transactions between different accounts are not compared and the anomalous sequences of events that may indicate fraudulent behavior can be identified.

### 2.3.3 Supervised techniques

**Bayesian Networks:** this is a method where by two networks constructed are used to describe the behavior of the user. Bayesian One network is constructed and it is then used to characterize a user as fraudulent (F) and Bayesian Two network is used to characterize a user as legitimate or non-fraudulent (NF). The fraud net is set up by using expert knowledge and the user net is set up by using data from non-fraudulent users. During operation, the user net is adapted by a specific user based on present data. By inserting evidence in the networks and propagating it through the network, the probability less than two is obtained. This shows at what degree the observed user behavior should meet typical fraudulent or non-fraudulent behavior.

Bayesian networks also allow the integration of expert knowledge, which is used for initial set up

in the models. On the other hand, the user model is retrained in an unsupervised way using data. Thus Bayesian approach incorporates both, expert knowledge and learning (Jayant, 2014).

**Hidden Markov Model:** Srivastava et al. (2008) defined HMM as “a double embedded stochastic process with two hierarchy levels”. It is used to model much more complicated stochastic processes. If a new transaction is not accepted with sufficiently high probability, it is considered to be a fraudulent transaction. Baum Welch algorithm is used for training purpose and K-means algorithm for clustering. In HMM, the data is stored in the form of clusters depending on three price value ranges low, medium and high. If the probabilities of initial set of transaction is chosen and then FDS will check whether transaction is genuine or fraudulent. Since HMM maintains a log for transactions it reduces the load of work on employees. However, it simultaneously produces high false alarm as well as high false positive. The initial choice of parameters which affects the performance of the algorithm should be chosen carefully (Sciences, 2013).

**Genetic algorithm:** One common and often recommended way of detecting fraud is the use of algorithms. Bentley et al. (2000) suggested an algorithm that is grounded on genetic programming so as to determine logic rules with the capability of classifying transactions from credit cards into their fraudulent and non-fraudulent classes. Genetic algorithms are inspired from natural evolution and they were first introduced by Holland (1975). Bentley et al. (2000) basically followed the scoring process. The database was made of 4,000 transactions with 62 fields from what was described in the experiment of their study. Training and testing models were engaged for the similarity tree. Various rules were tried with different fields. The one with the highest predictability was considered as the best rule.

Their technique demonstrated outcomes for genuine home protection information and one effective strategy of fraud detection in credit cards. Genetic algorithms are an evolutionary algorithm which provides better solutions as time progresses. Genetic algorithms are used in data mining mainly for variable selection and is mostly coupled with other data mining algorithms. Its combination with other techniques has a very good performance. Genetic algorithms are used in credit card fraud detection for reducing the wrongly classified number of transactions. And it is

easily accessible for computer programming language implementations which make it strong in credit card fraud detection. But this method has high performance and is quite expensive (Jayant, 2014).

**A Fusion Approach Using Dempster-Shafer Theory:** is where by information is fused and Bayesian learning in which evidences of current as well as past behaviors of an account are combined. An activity profile for every user is made depending on the certain shopping behavior. The system is made up of four components, rule-based filter, Dempster–Shafer adder, transaction history database and Bayesian learner. The transaction is classified as suspicious or non-suspicious depending on its initial stage. Once a transaction is found to be suspicious, belief is strengthened or weakened by comparing fraudulent or genuine transaction. Advantages of this approach are high accuracy, processing speed, reduces false alarm, improved detection rate and applicable in E-commerce. However, this approach is highly expensive.

**Neural networks:** these can be defined as a set of connected input/output units where each connection is represented by a certain weight. Networks learn by adjusting weights in order to predict correct class labels during the learning phase. An artificial neural network consists of an interconnected group of artificial neurons. The functions of the brain where pattern recognition is concerned and associative memory is the motivation behind the principle of neural networks. The neural network recognizes patterns which are similar, it predicts future values or events based upon the associative memory of the patterns it has learnt. It is widely applied in classification and clustering. There are two phases in neural network training and recognition. Learning is called training in neural networks. Neural networks can produce best result for only large transaction datasets hence they need a long training dataset.

**Decision tree algorithms:** These are rule based classifiers that utilize a divide and conquer method to construct prediction rules. The divide and conquer method works by breaking down a problem recursively into two or more sub-problems until it is simple enough to be solved directly. Decision trees are graphical representations of, ‘if-then’ statements. They consist of nodes and branches. The starting node is referred to as the root node and each node is labelled with a feature name and each branch leading out of it is also labelled but with one or more



possible values for that feature. Each node has only one incoming branch, except for the root, which is selected as the starting point. Each internal node in the tree corresponds to a test of the value of one of the features. Branches from the node are labelled with the possible values of the test and leaves are labelled with the values of the classification features and they specify the value to be returned if that leaf is reached.

Classification can take place as a result of traversing the tree by taking a set of features and their associated values as input. The feature of the instance corresponding to the label of the root of the tree is compared to the values on the root's outgoing branches, and the matching branch is selected. The process of node label matching and branch selection process continues until a terminal node, denoted to as leaf, is reached, and the case will be classified according to the label of the leaf and a decision made on the class assignment of the case (Pun, 2011). However, Decision trees are not stable and a small fluctuation in the data can make a large difference.

**Clustering:** is a whereby items are divided into significant groups which are alluded to as clusters. The items in a cluster alike to each other and not the same as items of another group. Clustering is otherwise called data segmentation or partitioning. It is recommended that data objects in each cluster ought to have high intra-group likeness inside that same cluster yet ought to have low likeness between groups of different clusters. A standout amongst the most well-known clustering systems are the K-nearest neighbor, the Naïve Bayes strategy and self-organizing maps.

## **2.4 Existing research about Credit Card Fraud Detection systems**

Fraud detection in credit cards has gotten a critical consideration from scientists globally. Several methods have been produced to recognize fraud on credit cards depending on neural system, genetic algorithms, clustering techniques, information mining, decision tree, Bayesian systems and so forth.

Ghosh and Reilly in 1994 proposed a neural system technique to distinguish transactions on credit cards. They built a system, which was prepared on an expansive example of credit card

transactions that were labelled. These specimens contained cases of stolen cards, lost cards, application extortion, stolen card subtle elements, fake misrepresentation and so forth. They tried on a dataset for a long period of time. (Vats, Dubey, and Pandey, 2013).

Also, Wiese et al. (2009) suggested the use of Artificial Neural Networks in credit card fraud detection. Their method considered a sequence of transactions that were of the past and used them to determine if the next transaction was legitimate or fraudulent. They believed that “looking at individual transactions” only was misleading since it could not air any periodical changes in spending behavior of a customer. They referred to their approach as the “Long Short-term Memory Recurrent Neural Network (LSTM).”

The problem with the discussed approaches with regard to neural networks is with the number of parameters, they have to be set before training begins. However, there are no rules for setting of the parameters yet they determine the success of the training. Networks differ in the way that their neurons are interconnected, in the way the output of a neuron is determined by its input (propagation function) and in their temporal behavior (synchronous, asynchronous or continuous). And the topology of a network has a large influence on the performance of that network but so far, no method exists to determine the optimal topology for a given problem because of the high complexity of large networks hence the choice of the basic parameters determine the success of the training process.

Also, there is the clustering technique which was proposed by Bolton and Hand in (2002). In this technique, clustering of two algorithms were used for detection of behavioral fraud. The system which was proposed identified accounts which were behaving differently from the rest at a given time whereas they were behaving the same prior by making use of peer group analysis. Such accounts were rendered suspicious and analysis took place for fraud. However, their technique did not work well with new customers.

An innovative implementation of SVMs, Support Vector Machines which is a binary classification technique for detecting credit card fraud was suggested by Rongchang Chen et al. (2004). It was based on the either of 0 or 1 i.e. genuine or fraudulent. They made a suggestion for

the credit card issuing companies and banks to ask new customers to fill in questionnaires to help the former in understanding spending patterns as well as habits of the customers. This approach proved to be very useful since there was no previous data on any customer and detection techniques could not spot fraudulent transactions at the initial stage. Answers to the questionnaires were therefore used in a similar way to the historical information of each customer.

They referred to their approach as the “Questionnaire-Responded Transaction Model” (QRT Model). The gathered Questionnaire-Responded Transaction (QRT) data from the online system was then considered as the transaction record and it was utilized to build up a personalized model, which was sequentially used to predict legitimacy of a new incoming transaction. Since the illegal user’s consumer behavior is usually dissimilar to the cardholder, fraud could be avoided from initial use of a credit card, even without any transaction data. However, one important issue regarding the QRT approach was how to predict accurately with only few data, say 100 to 200, since the users were usually not willing to answer too many questions.

(Weston, Hand, Adams, & Whitrow, 2008) suggested an implementation of PGA for detecting credit card fraud. However, their approach could not detect fraud in real time but once every night.

(Science, 2000) suggested a system which was based on boosted decision tree algorithms. This was an improvement on decision tree algorithms which were less stable especially given a slight change in the data. This approach was for generated and combined multiple classifiers, either decision trees or rule sets. This technique was used to improve the prediction accuracy of the classifiers. Instead of one classifier, several classifiers were constructed and the combination of their outcomes determined the final class being assigned to the case in boosted decision trees. The disadvantage of this approach was that the system was not adaptive hence the fraud environment is dynamic.

Srivastava et al. (2010) suggested an implementation of Hidden Markov Model which promised a good predictive accuracy and a minimal misclassification error. Their approach however did

not perform well with new customers since there was no prior historical information.

(Prakash & Chandrasekar, 2015) introduced an Optimized Multiple Semi-Hidden Markov Model, it was used to model parameters, the main aim of the system was to detect fraudulent users and optimize training values, and for optimization they made use of a Cuckoo search algorithm. Their main intent was on liberating customers from the necessity of statistical knowledge. They gathered information about the cardholder, observation symbols regarding to a particular cardholder was identified dynamically. The clustering algorithm then executed on past transactions, only the amount was taken for the cardholder. However, the main drawback was that it could not control or detect fraud before a real transaction was made, it focused on evaluating a transaction after it had already taken place meaning that it did not work in real-time.

(Rathore, 2016) suggested a hybrid technique for credit card fraud detection. It used the properties of Parallel Granular Neural Networks (PGNN) and Cost Based Model which was to help in detecting fraud acutely. With PGNN they could extract data from the database and also increase the speed of data extraction. They used a Fuzzy neural network based technique in PGNN whereby log data of different credit card transactions were used to detect fraud. Updated log data was required to detect fraud. In Cost Based Model a large dataset containing various credit card fraud events was presented. It was divided into smaller subsets, meta-classifiers, generated by the use of the Fuzzy neural network, the Cost Based Model was then integrated with the Fuzzy neural network to provide an efficient functionality and better performance for the fraud detection system. However, there was no automation of distributed data which degraded the performance of the whole technique. And also, there was no use enhanced security mechanisms i.e. the use of passwords for authentication of any transactions.

(Khandare, 2016) suggested a Hidden Markov Model to facilitate in stopping online transaction and off-line transaction through credit card. Using the Hidden Markov Model, the fraud detection processing system was trained with standard procedures and a user's spending patterns. The user had to provide information which included the credit card number, expiry month and year of credit card etc. If the transactions were less than ten the credit card fraud detection system checked if personal information was provided, after ten transactions it then started its

work. Where users failed any of the questions the transaction became unsuccessful.

However, its main disadvantage was that users were not always willing to offer too much personal information and to go through a lot of question just to purchase a simple item, hence the need of a dataset with past transactions to classify a transaction as it came and avoid the tedious job of answering too many questions was needed.

## **2.5 Inconsistencies and other shortcomings in our knowledge and understanding**

The advanced approaches that are used in credit card fraud detection include mostly neural networks, data mining, AI, inherent algorithm, support vector machine, etc. Most of the fraud detection systems as discussed above have a variation in their precision. The main issue is that most of the transactions that are identified as fraudulent by the systems are in fact legit. This in turn cause unnecessary losses to banks through investigations of the legit cases and it causes user frustration. Also, some most of the systems can only detect fraud after it has already taken place. Some of them lack adaptiveness as well.

## **2.6 Reasons for studying the problem further**

Almost all the existing fraud detection techniques try to capture and analyze behavioral patterns of users by checking for changes in subsequent transactions. However, these rules are largely static in nature, therefore false alarms are given if a user is to change his or her spending pattern. This in turn brands the techniques ineffective. The objective is to take in the behavior of clients progressively in order to limit loss. In this way, systems that can't advance or learn, may soon end up plainly obsolete bringing about expansive number of false alerts. A fraudster can in like manner indulge in new sorts of attacks which ought to be observed by the fraud detection systems.

For example, a fraudster may go for deciding most noteworthy favorable position either by making a few high purchases or endless low purchases to evade getting caught. Along these lines, there is a necessity for making systems which can organize different evidences including

cases of genuine examples from cardholders and that of fraudsters.

## **2.7 Contribution that the present study expected to make**

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. An application form for the user to fill in with the required information, passwords and range of income, location, occupation, address, gender, pin, email address among other details.
- ii. It should verify all the transactions making use of the Support Vector Machine which is a binary classification algorithm, which classifies transactions as either 1 or 0 i.e. fraudulent or genuine, which makes it easy to separate transactions and it also removes any unnecessary speculations or maybes on transactions.
- iii. The system should allow a purchase if the transaction is not fraudulent.
- iv. In the event that a transaction is suspicious: the system should prompt the user for his/her pin that he/she chose and is known only to him/her. If the pin is wrong the account closes and user would have to contact the bank and only one chance should be given to avoid trial and error or brute force attacks given that the client might be a fraudster.

Since most users are not always willing to share too much information from literature review only a few necessary details would be required from the user that is their address, data of birth, location, range of income, password, pin, gender as well as occupation. Transactions from new users are difficult to classify with no previous data in their accounts so those details would be used to classify their transactions.

In addition, datasets are not easy to obtain given that information of clients is confidential, so the researcher will synthetically create transactions. Also, a shop with different items, their categories either machinery, groceries and general upkeep, their locations, as well as their prices and quantity, using these the system should cater for new users correctly as well as old users to help in detecting any forms of fraud making use of the product's quantity, category, prize as well as its location if (iv) happens. For geographical location, the user made use of the ten provinces in Zimbabwe since it is a local based system. All this should happen before a transaction is

processed to detect fraud in real-time and also as quick as possible.

## **2.8 Chapter Summary**

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: Methodology**

### **3.0 Introduction**

The purpose of this study is to elaborate on how the research was done so as to fulfil the objectives of the research. (Rajaseker, 2010) postulated that a research methodology is a systematic way to solve a problem. It is a science of studying how research is to be carried out. Essentially, the processes by which scholars go about their work of defining, clarifying and forecasting phenomena are called research methodology. It is also defined as the study of methods by which knowledge is gained. Its aim is to give the work plan of research. In as much, the researcher in this chapter will also explain the choice of the methods used and an in-depth review of the design process of the whole research study giving a step-by-step outline of the design procedures as they were implemented in coming up with the desired system.

### **3.1 Research Design**

The design stage involves coming up with different modules of the system and their intended functionality. It is the process of defining the architecture, components, modules, interfaces and data for the system in-order to satisfy the requirements specified. The main objective is to make sure that an efficient, effective as well as reliable system is built.

The system interfaces should be designed with the user in mind so that they do not have any trouble or issues following through. The detailed system should be able to describe the data as well as information flow in the system and how the system works overally.



## **3.2 Data Collection Method**

### **3.2.0 Quantitative Approach**

For data collection, the researcher used a quantitative approach and this refers to information with amounts, qualities or numbers, making them quantifiable. In this manner, they are typically communicated in numerical form, for example, length, sum, size, cost, and even term. The utilization of understandings to produce and therefore break down this kind of information add trustworthiness or validity to it, so that quantitative information is generally seen as more dependable and objective.

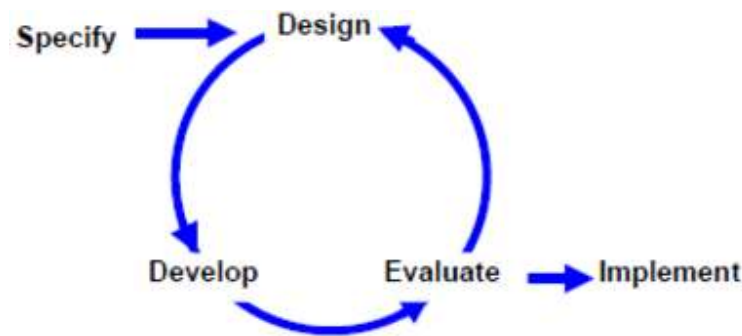
Due to the fact that datasets are not readily available to researchers, the researcher for data collection, created twenty accounts and each account had five transactions each, making the transactions a total of hundred. The transactions were then equally split to avoid imbalance into fifty for training and another fifty for testing. An experiment was then run on the fifty transactions for testing. For the two classes of training and testing samples they all held balanced suspicious and non- suspicious transaction respectively.

## **3.3 System Development Methodology**

A system development methodology talks about to the framework that is used to plan, structure, and control the process of developing an information system (Centers for Medicare & Medicaid Services, 2008). The system was created using rapid prototyping, the reason being that the system was developed in a limited amount of time and changes had to be welcome along the development of the system.

The researcher made use of Rapid Application Development (RAD), this is a kind of a software development that does not invest a great deal of energy or assets on arranging but rather utilizes a strategy for prototyping to present the item. A prototype is an adaptation of the item that emulates what that genuine item will resemble, and it can complete similar functions. This takes into account a speedier yield of the component being made. Without a wealth of preplanning in

the development phase, the model can undoubtedly be adjusted to rapidly roll out improvements all through the testing stages. RAD approaches emphasize versatility and the need of modifying requirements in light of learning picked up as the venture advances (James Martin, 1991)



*Figure 2: RAD Diagram.*

### **3.4 Requirements Specifications**

This part goes into detail about how the system is expected to respond to certain parameters and situations as illustrated below under functional as well as non-functional requirements.

#### **3.4.0 Functional requirements**

These can be defined as the function of a system or its components. A function in turn is a set of inputs, the behaviour, and outputs. Bittner stated that, “functional requirements are those actions that a system must be able to perform, without taking physical constraints into consideration. They might be computations, specialized subtle elements, data control and preparing and other particular functionalities that characterize what a system should achieve. The behavioural prerequisites portraying the vast majority of the situations where the system utilizes the functional requirements are illustrated in use cases.

In this research study, the system must be able to:

- allow new users to create accounts and old users to log in and purchase.
- capture, verify and classify transactions accordingly.
- authenticate details provided by the user and authorise purchases given a non-fraudulent transaction.
- block the account and tell the user to contact the bank if the pin is wrong and if the transaction has been rejected
- store all transactions in the database so that learning can continuously take place.
- give only one chance for the pin in account confirmation, so as to avoid brute force attacks from fraudsters and this in turn gives users a good interaction with their service providers as well as a sense of security.

### **3.4.1 Non-functional requirements**

Non-functional requirement can be defined as an elaboration on the performance characteristic of a particular system. They describe how well or to what standard a function should be provided e.g. levels of required service such as response times, security and access requirements, usability, performance supportability as well as project constraints such as implementation on the organisation's hardware/software platform among other factors. The most important of all the non-functional requirements is the ability of the system to provide testability and maintainability. 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:

- be up and working clock round.
- make sure that transactions are processed as fast as possible to avoid user frustration. And give feedback.

## **3.5 Design Tools**

### **3.5.0 Hardware design tools**

Laptop- this has been used in the development of the system.

### **3.5.1 Software design tools**

These can be defined as objects, media, or computer programs that are used in the development of a system. They influence the process of production, expression and perception of design ideas and therefore they need to be applied skilfully so as to produce the best as well as the desired results. The used the tools below in the development of the system:

- PHP language
- MySQL
- sublime text

#### **PHP Language**

It is a free language with no licensing fees so the cost of using it is minimal. A good benefit of using PHP is that it can interact with many different database languages including MySQL, it is also a free language so it makes sense to use PHP. It has a very good online documentation with a good framework of functions in place. This makes the language relatively easy to learn and very well supported online. PHP also codes runs much faster.

#### **MySQL Database**

This is a relational database management system (RDBMS) which was adopted for its typical use for web application development (often accessed using PHP); thus, making it the perfect candidate for use. In addition, it generally offers fewer features than other databases, which means that it is fast and due to its speed, the system will overall be fast.

## **Sublime text editor**

This is a branded code editor which is cross-platform and it can be used for programming as well as markup languages. It is a free-software license software. Its advantages include the ability to simultaneously edit code, to auto save as well as correction spells as one codes.

## **3.6 Description of the System**

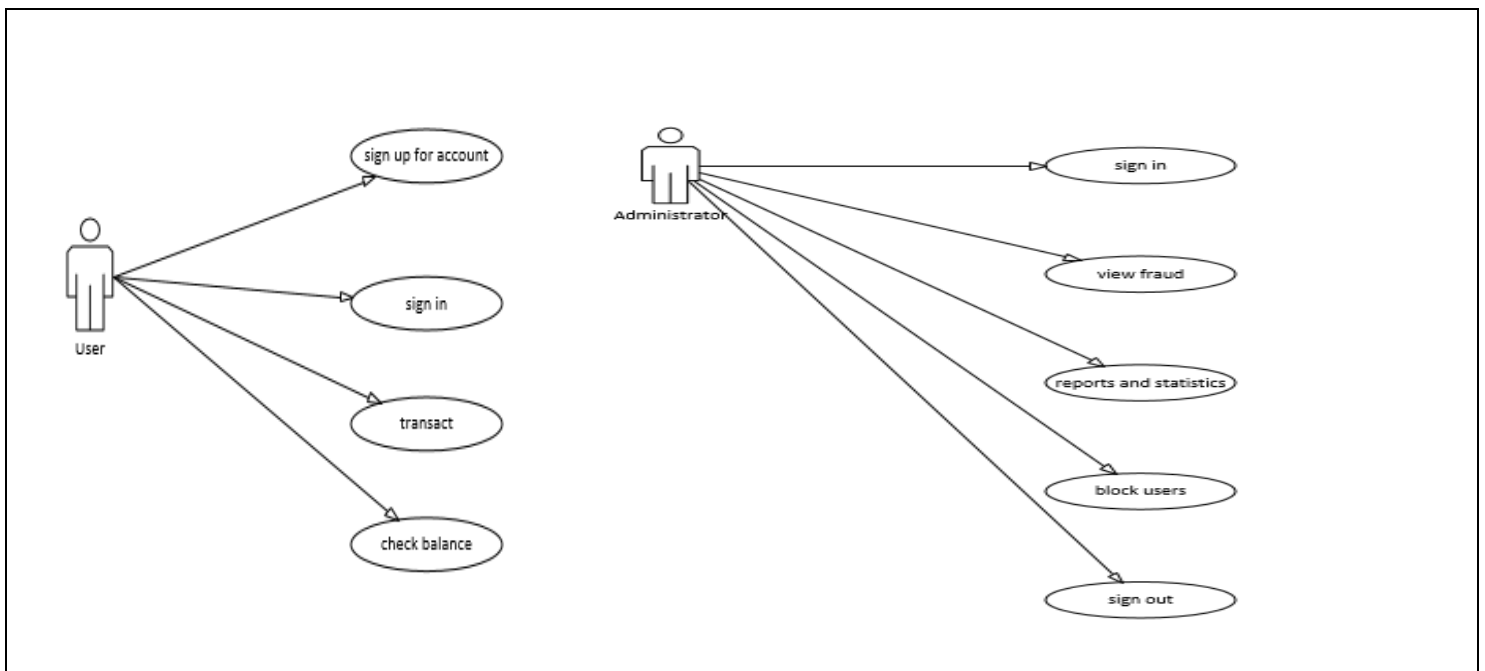
### **3.6.0 System main users**

- User
- Administrator

### 3.6.1 Use Case Diagrams

The author will utilize use case diagrams to illustrate the interaction of the user as well as the administrator with the system. A use case is a portrayal of a client's connection or interaction with the system that demonstrates the connection between the user and the diverse use cases in which the user is included in, its function is also to capture the dynamic aspect of a system in general. So as to define the functional requirements of the research study use case diagrams are going to be used for the whole process of the development.

The new user can create an account, get a credit card number, sign in and transact. An administrator can also sign in, do necessary changes as illustrated below.



*Figure 3: System Use Cases*

### 3.6.2 System flow chart for Supervised Learning.

Below is a data flow diagram that illustrates the processes 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 algorithm on that training set. The dataset is then split into two thus for training the algorithm and testing the system thereafter an algorithm is implemented, in this research the author made use of the Support Vector Machine which is a binary classification algorithm.

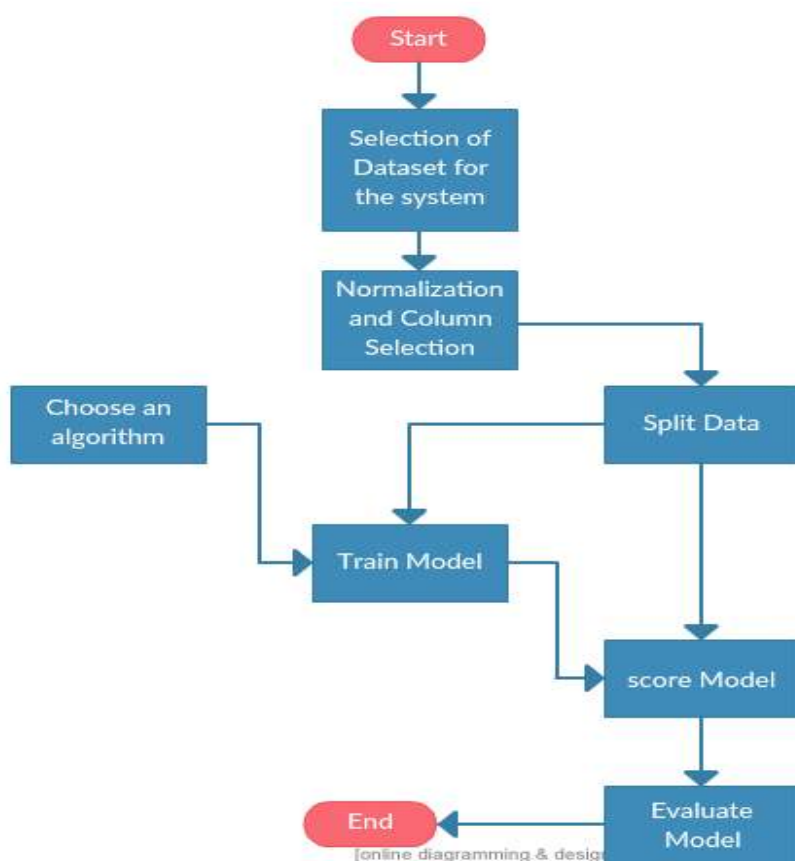


Figure 4: Supervised learning flow chart

### 3.6.3 System flow chart for Credit Card Fraud detection using Supervised Learning

From the diagram given below a new user provides the required details which are captured into the system and stored in the MySQL database for use, he/she is granted a credit card number as well as a verification card number. Thereafter a user can sign in and transact as desired within the specified ranges of his/her card limit and income. All the transactions are tested for legitimacy by the system, if a particular transaction is legitimate then user can proceed to purchase otherwise user has to provide a pin. If pin is wrong or user has forgotten he/she must contact bank as well as if transaction fails to process. Moreover, transactions are recorded in the database for adaptive learning of the Fraud detection system. See fig below.

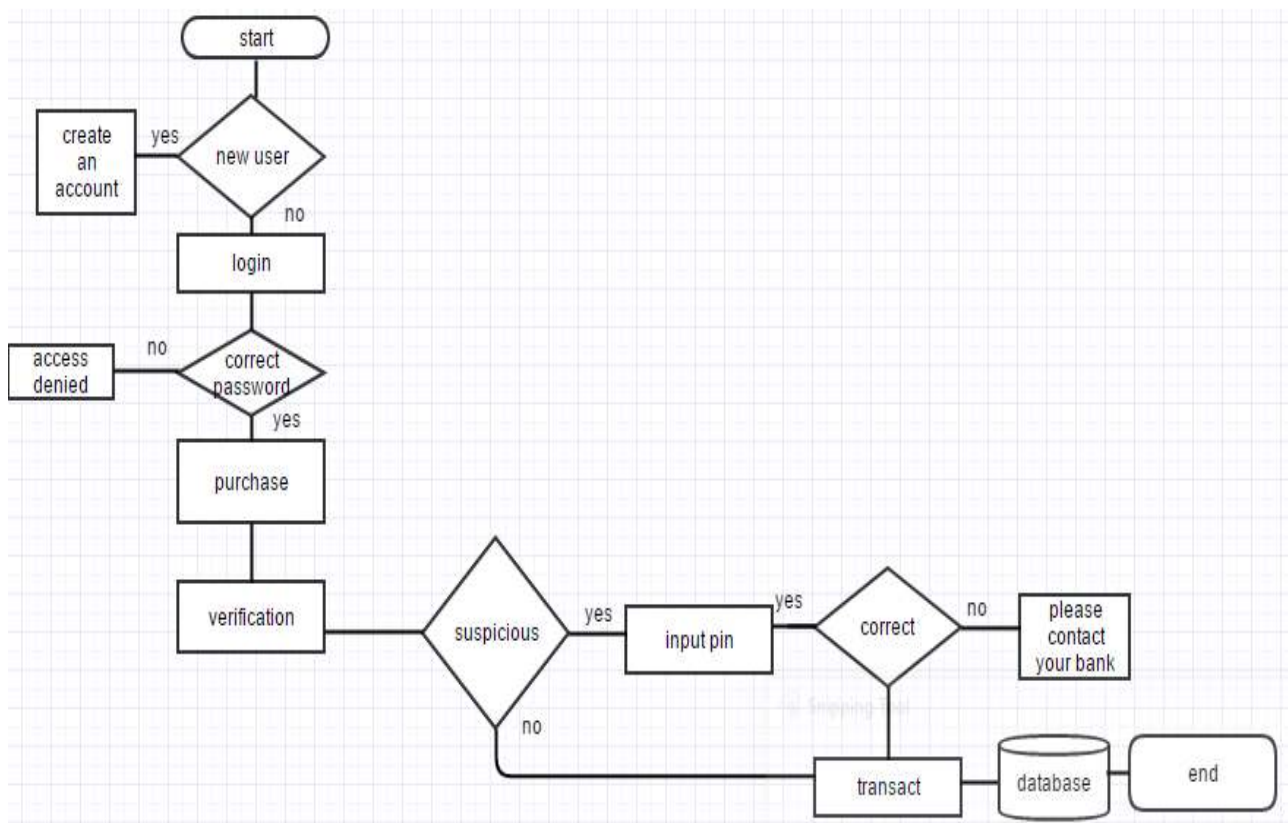


Figure 5: Overall system flow chart.



### 3.6.4 The Relational Database Schema

The author highlighted a bit above on the choice of a database thus the MySQL Server which is a relational database adopted to work with the credit card fraud detection system. Some of the reasons for this choice include the ability to hold data in separate tables thereby making it easy to add records. Also the database tasks take little time to launch, leaving one to focus on the result. A shop was created by the user with different products of different values as well as different categories and location. Transactions that were created were stored in the database, also the user accounts as shown below on the ER diagram.

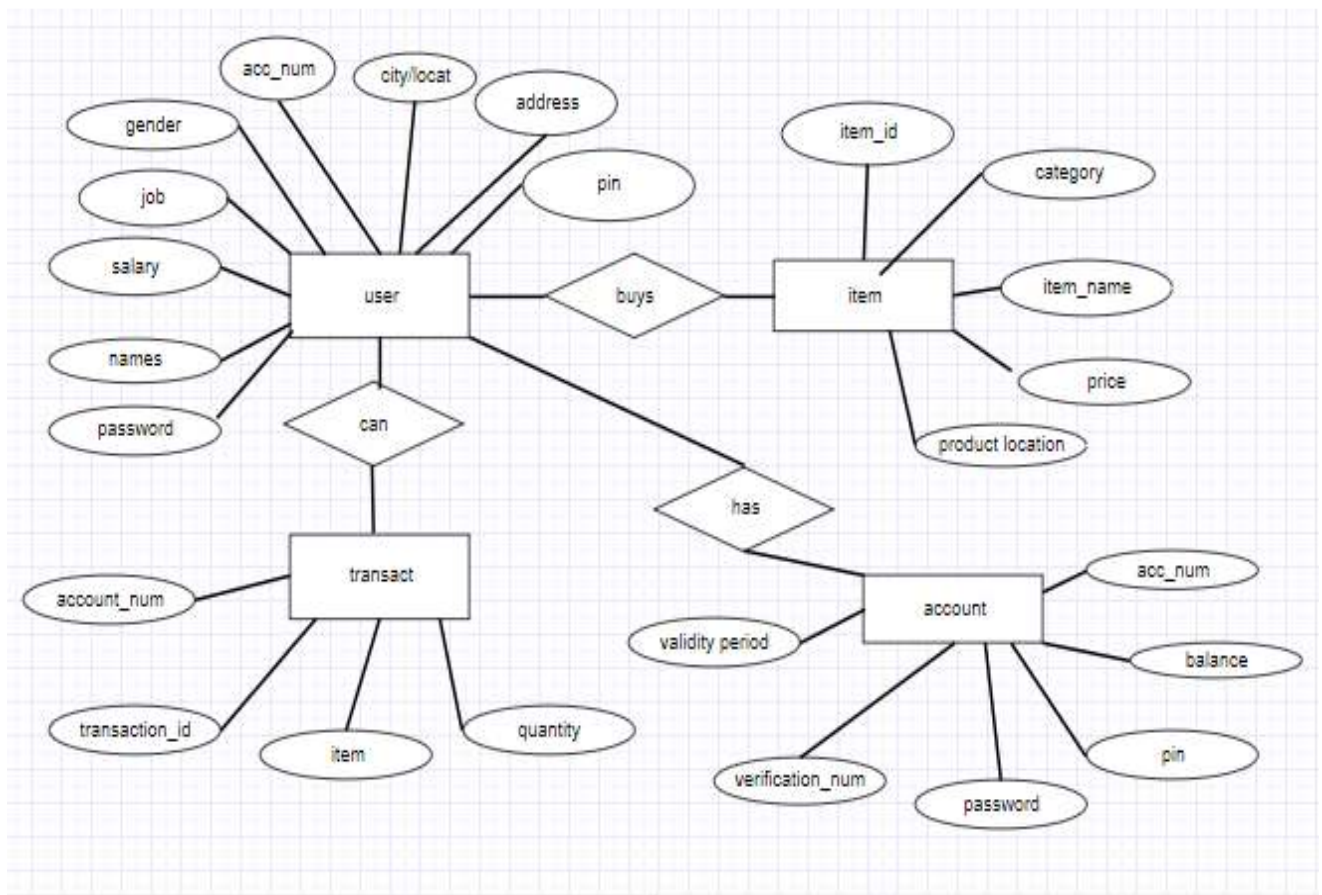


Figure 6: ER diagram.

## **3.7 Software Design**

### **3.7.0 User Interface Design**

In-order to come up with the best and most interactive user interface there is a need for a software development model that would be able to integrate with the required components. A user interface design has to do with giving a description of a user's behaviour (Jess James Garrett, 2011). It can also be described as a component of a computer system that which the user interacts with in-order to perform the desired tasks. Some of the factors considered by the author were:

➤ **Simplicity**

The user interface was designed in such a way that users, even naive ones can be able to access everything they want without any trouble given that everything is clearly laid out.

➤ **Size**

In-order to ensure lightness of the overall system hence improved speed on user requests the size was minimized, that of the user interface.

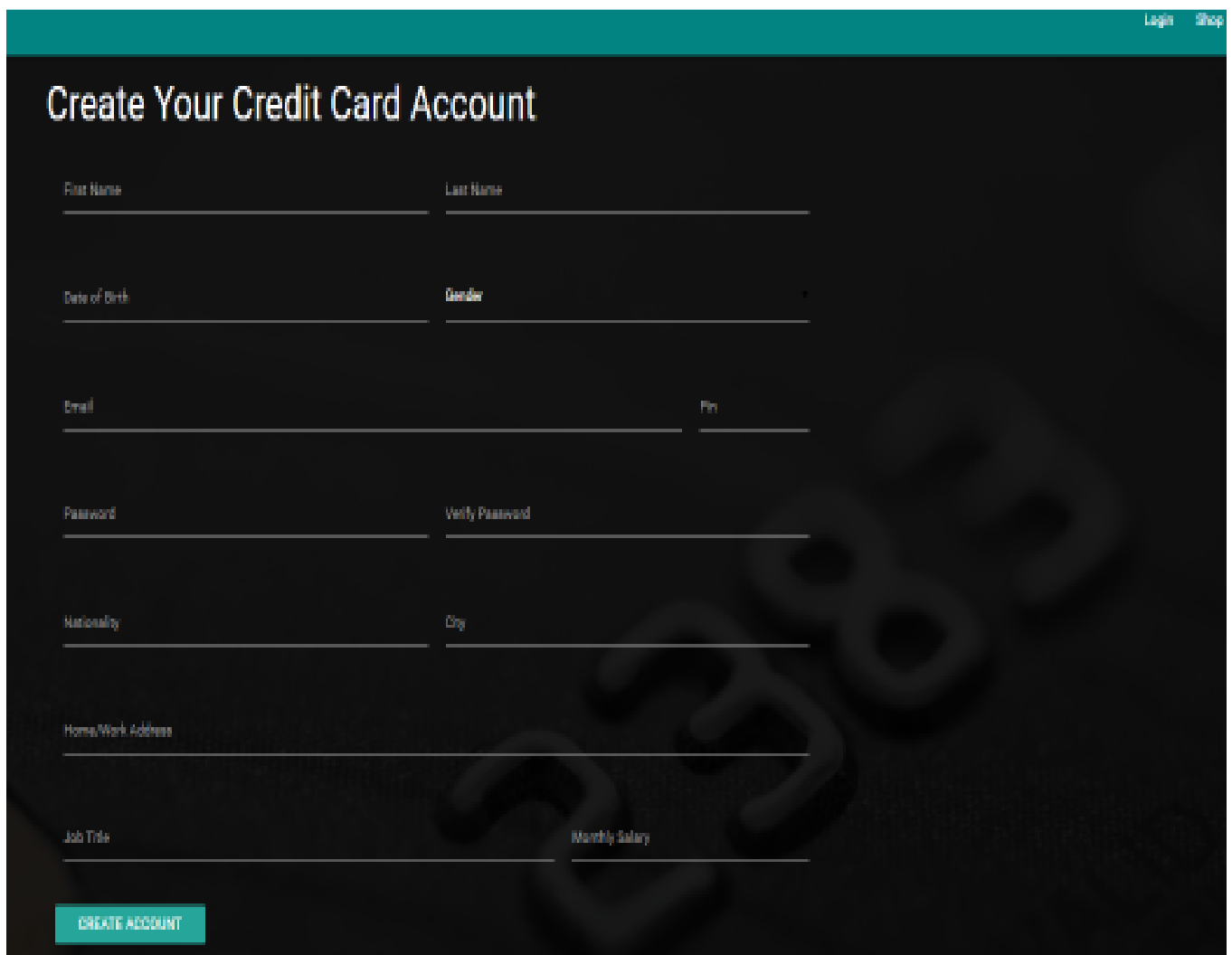
➤ **Structure**

The interface was built in a simple way so that users could navigate without making unnecessary mistakes or wasting time. Also, there are various options for the new users, clearly set.

### 3.7.1 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.

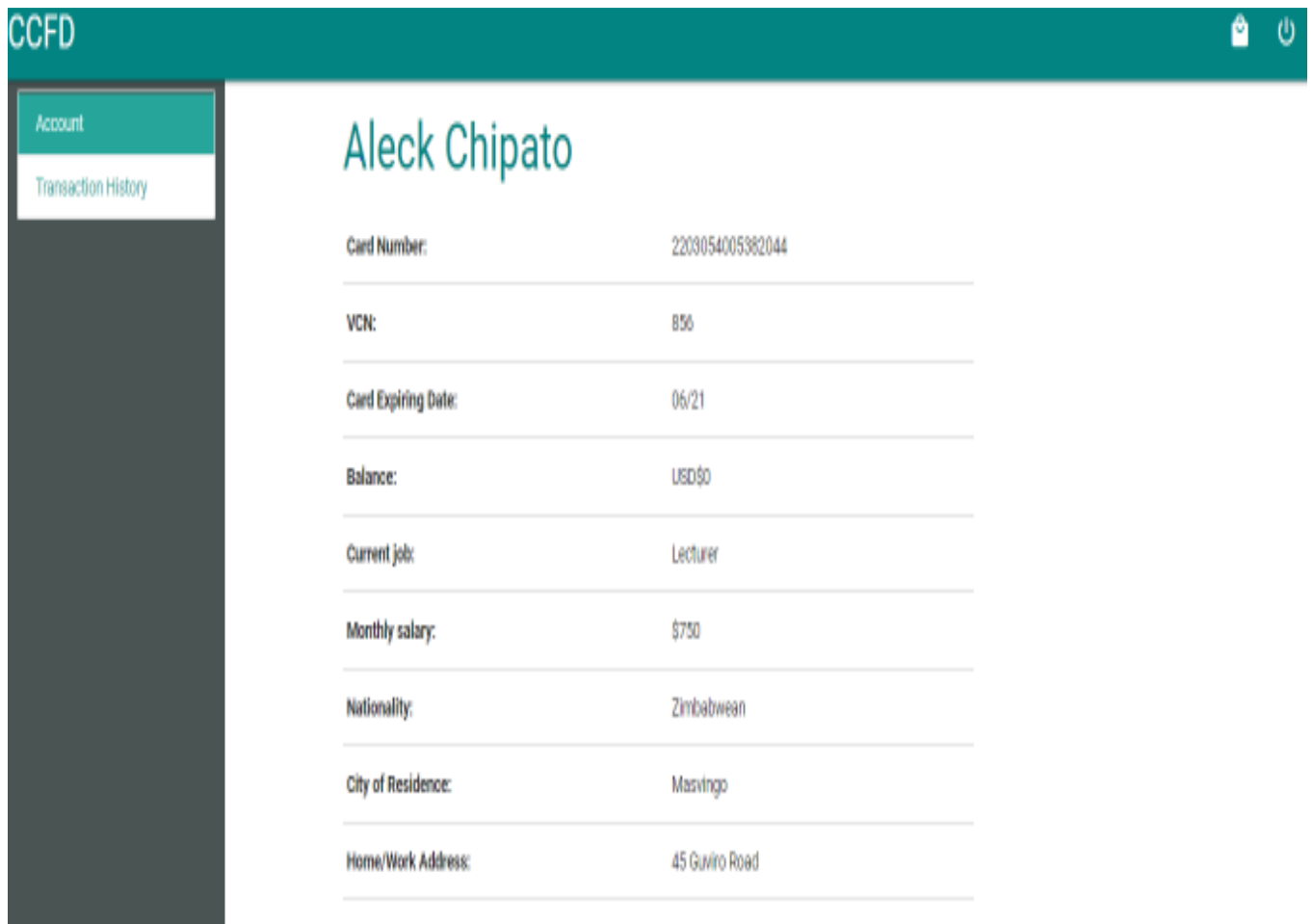
Below are screenshots showing the Registration and Login Page. User has to first apply for the desired card type, filling in the required details as shown below.



The screenshot displays a web interface for creating a credit card account. At the top right, there are links for 'Login' and 'Shop'. The main heading is 'Create Your Credit Card Account'. The form consists of several input fields arranged in a structured layout: 'First Name' and 'Last Name' (two columns), 'Date of Birth' and 'Gender' (two columns), 'Email' and 'Pin' (two columns), 'Password' and 'Verify Password' (two columns), 'Nationality' and 'City' (two columns), 'Home/Work Address' (one column), and 'Job Title' and 'Monthly Salary' (two columns). A teal 'CREATE ACCOUNT' button is located at the bottom left of the form area. A large, faint '2383' watermark is visible across the right side of the form.

*Figure 7: Account creation and login in platform.*

Below is a screenshot showing the details that a user can view, transaction history button, his/her remaining balance, address, Verification Card Number etc.



*Figure 8: Account history.*

Shown below is a diagram showing the shop with various products, category is shown on the drop-down button, in it are groceries, general upkeep as well as gadgets/machinery sections in the shop, locations are given together with the prices of the products as shown below. A user can select their category as well as city to buy from.

Best Shop

Category

Location

Harare

Bulawayo

Chinhoyi

Marvingo

Bindura

Mutare

Gweru

Marondera

HP Laptop	Harare	Gadgets/Machinery	\$600	BUY
Kettle	Harare	Gadgets/Machinery	\$20	BUY
stove	Harare	Gadgets/Machinery	\$500	BUY
car	Harare	Gadgets/Machinery	\$7500	BUY
sofas	Harare	Gadgets/Machinery	\$700	BUY
kitchen chairs	Harare	Gadgets/Machinery	\$500	BUY
tv	Harare	Gadgets/Machinery	\$550	BUY
fridge	Harare	Gadgets/Machinery	\$900	BUY

Figure 9: Online Shopping mall diagram

Given below is a snapshot of the administrator's area of operations where he/she can delete, modify and make any changes that he/she sees necessary. The administrator can also view card types of users, view fraud etc.



Figure 10: Administrator interface.

### 3.8 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 PHP, MySQL as well as the sublime text 3 illustrating in detail how they related with each other to come up with a functional system. Functional and non-functional requirements definitions were given in relation to the development of the system. In addition, interactions between the system and user as well as the relational database have been given together with the Entity Relationship (ER) diagrams and dataflow diagrams. All the aspects mentioned earlier on in the scope of the project were covered so as to fulfil the aims and objectives of the project. The results are going to be shown in chapter four.

## **Chapter Four: Data Analysis Presentation, Analysis and Interpretation**

### **4.0 Introduction**

This chapter gives a summary of what the researcher's finding where in the form of tables, graphs as well as statistics. The researcher will use analysis to conclude the relevance of the system. The analysis is going to be used by the researcher to conclude on the system's relevance. A sample of 50 transactions were used for analysis. Microsoft excel was used in displaying the research findings.

### **4.1 Data Analysis**

Fraud analysis has the legal and the positive/fraudulent class and they are sub-divided into the following.

P – positive transactions

N – legal transactions

TP – legal transactions projected as legal

TN – fraudulent transactions projected as fraudulent

FN – fraudulent transactions projected as legal

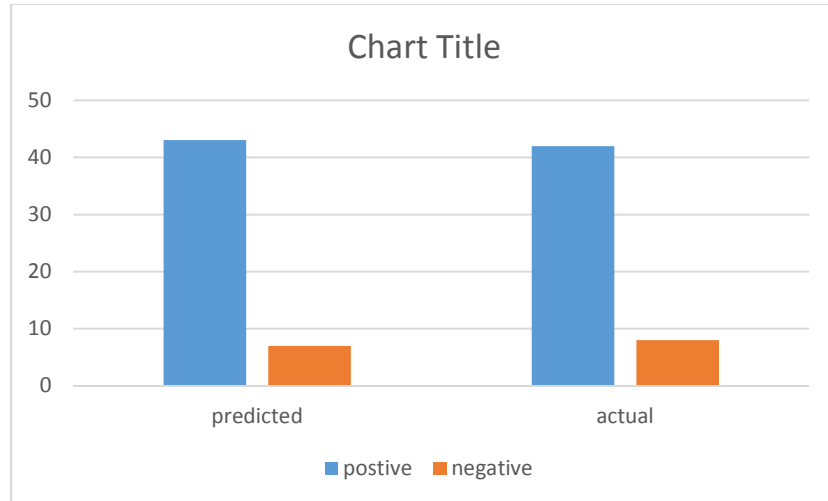
FP – legal transactions projected as fraudulent

The researcher used a confusion matrix and it has, columns denoting predicted classes and rows denoting the actual classes. It is a table that is usually used to define the performance given by a classification model on test data where by the true values are known. From it accuracy, error, recall and precision were calculated.

*Table 1: Confusion Matrix*

Actual	N=50	Predicted		
		Positive	Negative	
	Positive	TP =40	FN =2	42
	Negative	FP =3	TN=5	8
Total		43	7	50

Below is a graph illustrating the confusion matrix.



*Figure 11: Confusion Matrix Graph*

The researcher went on to calculate the accuracy and error rate of the system, illustrating them statistically with a bar graph.



#### 4.1.0 Accuracy Rate

Accuracy is defined as the most intuitive performance measure and it can also be characterised as the level of closeness of the anticipated value to the real value. It is given by:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+RFP+FN)}$$

$$= (45/50) * 100$$

$$= 0.90$$

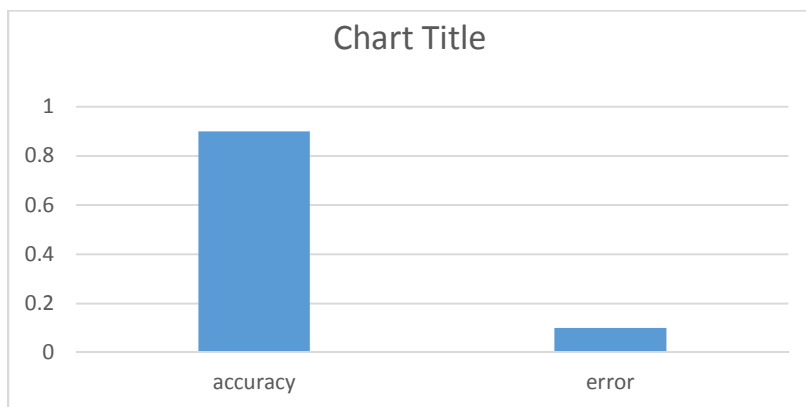
#### 4.1.1 Error Rate

Error rate is defined as an assessment of the likelihood of an error happening during the end of a task or rather as a measurement of the efficiency of a system. It is also the ratio of the sum of erroneous units of data to sum of units of data conducted.

$$\text{Error} = \frac{(FP+FN)}{(TP+FP+FN+TN)}$$

$$= (5/50) * 100$$

$$= 0.10$$



*Figure 12: Accuracy and Error rate diagram.*

Moreover, recall and precision was calculated by the researcher as below.

#### **4.1.2 Recall Rate**

Recall is characterised as genuine positives (TP) divided by the aggregate number of components that belong to that class of genuine positives and false negatives.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$= 40 / (40 + 2)$$

$$= 0.95$$

#### **4.1.3 Precision Rate**

Is defined as elements correctly labelled as belonging to true positives divided by the sum of elements predicted as belonging to true positives and false positives. High precision does relate to the low false positive rate.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$= 40 / (40 + 3)$$

$$= 0.93$$

From the calculations made accuracy is inverse to error, the lower the error the higher the accuracy. Also, high fraud detection rate and low false rate are obtained by choosing appropriate parameter values obtained by checking spending patterns of cardholders. The model gave more than 80 percent accuracy which is more important. This shows that the system was successful in in classifying transactions correctly to a greater extend with only a low percentage of error rate for those transactions that were wrongly classified. Achieving high accuracy is vital and reducing the false alarms are also important tasks in the credit card fraud detection. Too much false alarms annoy users and must be reduced, this reduction doesn't affect accuracy though.

In addition, precision and recall was calculated giving percentages above ninety percent signifying that the system performed very well in correctly classifying fraudulent as well as non-fraudulent transactions.

## **4.2 Software Testing**

Software testing is whereby a program is run with the sole intent of finding errors. The main goal is to convince those in need of the system that the system is what they expected and is good enough for use. It is intended for building confidence in the system, so the activities are planned first and done methodically. The process of testing a software is talked about as verification and validation.

The researcher used various types of testing on the system which will be illustrated below:

### **4.2.0 Black box Testing**

Black box testing is also referred to as behavioral testing. This type of testing is more concerned with functional requirements of the program, it allows derivation of a set of input conditions to exercise all functional requirements. The key resolution is to check if the system is working to anticipations and meeting user expectations also.

#### **4.2.1 Fuzzy Testing**

Fuzz testing is one of the black box testing techniques. This is done by testing actions with random data so as to perceive any bugs in the system and weaknesses where input is concerned. This technique was used by the researcher and there were no errors.

#### **4.2.2 Alpha Testing**

This is an acceptance testing technique that is carried out before the project is done or release for use so as to make essential design changes. This was carried out by the researcher and the supervisor. Problems were noted down and rectified.

### **4.2.3 Beta Testing**

This is considered as the second phase in testing a system and it is done after the system is complete by real users of the system. This is done so as to test the functionality as well as usability of the system. This is usually the final phase before release to customers. This phase was conducted by the supervisor as well as some of the students.

### **4.2.4 Performance Testing**

This is a type of a non-functional testing technique and it is done to conclude the speed of the system's execution among other performance features. The system's performance was tested on the database in terms of retrieval and manipulation, also to test the speed of classification of the system. Satisfactory results were obtained.

## **4.3 Chapter Conclusion**

Based on the results observed by the researcher it is evident that the system can satisfy the H1 hypothesis which states that there will be success in the detection of credit card fraud. However, testing the system was a very difficult task since there was no similar data to compare to. This is due to the fact banks do not share their data with researchers and for those who had the privilege to that data only statistics and graphs are there to show for it.

However, using the Support Vector Machine algorithm the accuracy rate was improved given that it is rendered as medium, the accuracy, also the speed of classification was faster. This is due to the fact that the algorithm was working with a dataset that had only the necessary information, there was no overload of unnecessary details. Also, the dataset did not have hundreds of transactions that would take too long to process and predict on since the algorithm works best with a smaller dataset. The next chapter will take a look at recommendations, future work as well as challenges encountered.

## **Chapter Five: Conclusions and Recommendations**

### **5.0 Introduction**

This is the final chapter and it aims at the conclusions that can be drawn from the research and the suggestions that the researcher found from the results of the research. Furthermore, it investigates the future works that should be possible to make strides in the detection of fraud on ‘on-line’ systems. The outcome of this chapter will give the organizations interested in providing credit card services an insight on the issues that would need to be addressed so as to limit the number of fraud cases. Generally, this chapter looks at those features which are important but were outside the scope of the author’s project.

### **5.1 Aims and Objective Realization**

The researcher’s aim was to come up with a credit card fraud detection system that is based on machine learning to be used by service providers of credit cards. The researcher developed a web based system. Basing on the results accumulated the scientist can securely say that the research objectives were accomplished to a more noteworthy extent. The research objectives were as follows:

- To examine on the various techniques of “on-line” 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, analyze and evaluate the developed credit card fraud detection.

The results from the prior chapter have clearly presented that the system can significantly detect and reduce credit card fraud. Therefore, to a large extent, the researcher concludes that the system managed to answer 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?

## **5.2 Challenges Encountered**

Various challenges were encountered by the researcher which include the availability of datasets for supervised learning. This is due to the fact the transactions are sensitive and organizations try by all means to protect their clients' privacy and from attacks. Also from the datasets that are available they have no labels in an effort to protect clients making it difficult to use them in any way and with heavy imbalances thus the number of fraudulent transactions is much smaller than legitimate ones. History shows that models trained on such data do not perform well hence a problem for fraud detection techniques. Also since there are not yet any functional credit card systems in local Zimbabwean banks the system could not be implemented for testing on a real platform in organizations.

## **5.3 Recommendations**

Where credit card companies are concerned, transactions continually change and users may change their preferences or change in their behavior such that the data that was collected in the beginning becomes invalid so there is a need for a database for each customer such that his/her spending behavior is not determined by an unknown dataset. The system can then learn and adapt to each new incoming transaction.

In addition, current credit card fraud detection techniques can only identify fraud after it has taken place more than once usually. So there is a need for more techniques which can detect fraud in real-time because by the time fraud is detected usually large amounts would have been stolen. To protect customers, it is vital for financial organizations to consider real-time detection systems which can prevent losses.

Lastly, banks are more worried about covering up any gaps in their transaction security. However, focus should be on detecting and preventing fraud by working together and sharing information with other organizations. Technology, regulations and fraudsters will all continue to evolve and an industrywide view of fraudulent activity will be fundamental in making in-roads into the ongoing fraud battle.

## **5.4 Future Work**

There are a few improvements that can be done to the system in the future, these include the need for a system that focuses not only on withdrawals for behavior analysis but deposits and paying back patterns.

Furthermore, there is need to consider all forms of fraud since online fraud is not the only form of fraud on credit cards.

Also, credit cards are not conformed to only a universal limit, limits and card types depend on an individual's income, so there is a need for a system which puts all that into consideration.

There is also a need to make use of SMS to notify users of transactions made and attempted to his/her account.

## **5.5 Conclusion**

The conclusion that can be drawn from the research is that the project was a success to a larger extend as it managed to fulfil the objectives outlined in Chapter One and the researcher accepts h1 in the first chapter to say that there was success in the detection of credit card fraud given the results and analysis.

## References

- Jayant, P. (2014). Survey on Credit Card Fraud Detection Techniques, 3(3), 1545–1551.
- Journal, I., & Trends, C. (2014). Analysis on Credit Card Fraud Detection Methods, 8(1), 45–51.
- Khandare, N. B. (2016). Credit Card Fraud Detection Using Hidden Markov Model, 1(4), 83–86.
- Prakash, A., & Chandrasekar, C. (2015). An Optimized Multiple Semi-Hidden Markov Model for Credit Card Fraud Detection, 8(January), 165–171.
- Pun, J. K. (2011). Improving Credit Card Fraud Detection using a Meta-Learning Strategy by.
- Rathore, S. (2016). A Hybrid Technique for Credit Card Fraud Detection, 5(5), 20–23.
- Review Paper on Credit Card Fraud Detection. (2013), 4(7).
- Science, A. (2000). The Enhancement of Credit Card Fraud Detection Systems using Machine Learning Methodology.
- Sciences, P. (2013). Design and Implementation of a Fraud Detection Expert System using Ontology- Based Techniques A dissertation submitted to the University of Manchester Giannis Potamitis School of Computer Science Table of Contents.
- Sharma, A. (2012). A Review of Financial Accounting Fraud Detection based on Data Mining Techniques, 39(1), 37–47.
- Vats, S., Dubey, S. K., & Pandey, N. K. (2013). A TOOL FOR EFFECTIVE DETECTION OF FRAUD IN CREDIT CARD SYSTEM, (1), 25–29.
- Weston, D. J., Hand, D. J., Adams, N. M., & Whitrow, C. (2008). Plastic card fraud detection using peer group analysis, 45–62. <https://doi.org/10.1007/s11634-008-0021-8>
- Oxford Concise English Dictionary, 11th Edition, Oxford University Press, 2009.
- A. Srivastava, A. Kundu, S. Sural and A. K. Majumdar, “Credit Card Fraud Detection Using Hidden Markov Model,” IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, vol. 5, no. 1, pp. 37-48, 2008.
- S. Hawkins, H. He, G. Williams and R. Baxter, “Outlier Detection Using Replicator Neural Networks,” CSIRO Mathematical and Information Sciences, Australia, 2002.



- B. Wiese and C. Omlin, "Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks," *Innovations in Neural Infor. Paradigms & Appli.*, pp. 231-268, 2009.
- T. GUO and G.-Y. LI, "NEURAL DATA MINING FOR CREDIT CARD FRAUD DETECTION," in *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics*, Kunming, 2008.
- M.-S. Chen, J. Han and P. S. Yu, "Data mining: An Overview from a Database Perspective," *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, vol. 8, no. 6, pp. 866-883, 1996
- R.-C. Chen, T.-S. Chen and C.-C. Lin, "A new binary support vector system for increasing detection rate of credit card fraud," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 20, no. 2, pp. 227-239, 2006.
- R.-C. Chen, M.-L. Chiu, Y.-L. Huang and L.-T. Chen, "Detecting Credit Card Fraud by Using Questionnaire-Responded Transaction Model Based on Support Vector Machines," *Springer-Verlag Berlin Heidelberg*, pp. 800-806, 2004.
- W.-F. YU and N. Wang, "Research on Credit Card Fraud Detection Model Based on Distance Sum," *International Joint Conference on Artificial Intelligence*, pp. 353-356, 2009.