UNIVERSITY OF GHANA, LEGON
SCHOOL OF BASIC AND APPLIED SCIENCE
DEPARTMENT OF STATISTICS & ACTUARIAL SCIENCE

UNIVERSITY OF GHANA

CREDIT CARD FRAUD DETECTION ; A MACHINE LEARNING APPROACH

BY

JOHN GLAH

(10702625)

NOVEMBER, 2020

UNIVERSITY OF GHANA, LEGON
SCHOOL OF BASIC AND APPLIED SCIENCE
DEPARTMENT OF STATISTICS & ACTUARIAL SCIENCE

UNIVERSITY OF GHANA

CREDIT CARD FRAUD DETECTION ; A MACHINE LEARNING APPROACH

BY

JOHN GLAH

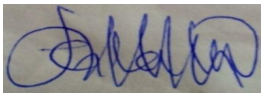
(10702625)

THIS THESIS IS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES, UNIVERSITY OF GHANA IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF THE MASTER OF PHILOSOPHY DEGREE IN STATISTICS.

NOVEMBER, 2020

Declaration
Student's Declaration

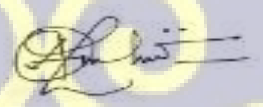
I hereby declare that this thesis work is as a result of my personal effort toward the award of Master of Philosophy Degree (MPhil.) and to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of this university or any other universities except where due acknowledgment has been made in the text.



JOHN GLAH ....03./06/2022..
Student ID (10702625) (Signature) (Date)

Supervisors' Certification

I hereby declare that the preparation and presentation of this thesis work was supervised in accordance with the guidelines of supervision on thesis laid down by the University of Ghana.



Certified by:
Dr. Louis Asiedu ....03/06/2022...
(Principal Supervisor) (Signature) (Date)

Certified by:
Dr. Samuel Iddi ....03/06/2022..
(Co-Supervisor) (Signature) (Date)

Dedication

I dedicate this honour to the Almighty God for bringing this work to fruition. Also to my daughter Ntifafa.

Acknowledgment

I owe it a duty to express my profound gratitude to all who in diverse ways assisted me to complete this academic exercise. First and foremost, I thank God the Father, the Son and the Holy Spirit for the divine inspiration, sufficient Grace and Strength without which nothing would have been done. I express my deepest gratitude to Dr Louis Asiedu and Dr Samuel Iddi my supervisors who guided me throughout this study. They supervised my work critically but with patience, encouragement, suggestions and did all necessary corrections without hesitation. Their readiness and willingness to give immediate attention under difficult circumstances and time constraints were invaluable to me. I feel most honoured to have had them as supervisors. My sincere gratitude goes to all the lecturers in the department. I thank Mr Terrison Adeho Yao Francis for his immense financial support throughout the duration of the programme. May God richly bless you. Finally, I thank my family and friends who supported me in divers ways to finish the programme. I thank my beloved wife Mabel for her encouraging words, and all my course mates for their help and support especially Bernard Oboh Essah and Diana Ayorkor Agbenyega.

Abbreviations

ATM	Automated Teller Machine
RBF	Radial Basis Function
RCE	Restricted Coulomb Energy
FPM	Fraud Pattern Mining
SMOTE	Synthetic Minority Over - Sampling Technique
ANN	Artificial Neural Network
NFMS	Neural Fraud Management Systems
SVM	Support Vector Machine
FMS	Fraud Management Systems
BP	Back Propagation
PIN	Personal Identification Number
AIS	Artificial Immune Systems
QRT	Questionnaire Responded Transaction
FDS	Fraud Detection Systems
HMM	Hidden Markov Model
IDS	Intrusion Detection Systems
QP	Quadratic Programming
KNN	K - Nearest Neighbor
PCA	Principal Component Analysis
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

Abstract

In recent times, credit card usage has increased tremendously because it is convenient to use and also saves a lot of time. Credit cards are rectangular plastic cards issued by banks which allow a person to borrow funds from a pre - approved limit to pay for one's purchases now and pay later. In the same manner, credit card frauds have also been on the increase causing huge sums of financial loss to credit card issuers. Credit card fraud is the use of a credit card by someone who is not the owner of the card and is not allowed to use it. In this study, three classification methods were used to do a deep analysis of credit card transactions history and the fraud detection models built. This study presents and demonstrates the advantages of support vector machine, artificial neural network and the k - nearest neighbor algorithms to the credit cards data for the purpose of reducing the bank's losses. The results show that the linear support vector machine and k - nearest neighbor approaches outperform artificial neural network in solving the problem under investigation. This study allows for multiple algorithms to be integrated together as modules and their results combined to increase the accuracy of the final results.

Contents

List of Tables

List of Figures

Chapter One

Introduction

1.1 Background of the Study

Data mining is about extracting understandings from data that must be statistically accurate, known before and with good purpose Elkan (2001). The data should be accessible, important, sufficient and clean. Challenges of data mining should be very well explained, can not be overcome by doubting and reporting instruments, and is motivated by models of data mining processes Lavrac et al. (2004). Fraud refers to exploiting the structure of a profit corporation without actually contributing to direct legal issues. Fraud can become a critical issue for an organization in a competitive market if it is very common and if the preventive mechanisms are not effective enough. Fraud detection automates and helps to reduce the manual components of a screening procedure, which is part of the general fraud control systems. This sector has become one of the most developed data mining applications in industry and government. Rationale and the intent behind a transaction can not be absolutely assured. In reality, the most cost - effective approach will be to use statistical algorithms to search for potential proof of fraud from

the available data. The analytical engines within these solutions are developed from various research communities, particularly from developed countries where software is regulated using artificial immune systems, artificial intelligence, auditing, databases, distributed and parallel computing, econometrics, expert systems, fuzzy logic, genetic algorithms, machine learning, neural networks,recognition of patterns, figures, visualization and others. Several advanced solutions and tools are used for fraud detection that safeguard credit card issuers. There exists two major criticisms of research on data mining - based fraud detection: 1. the limited amount of real life data accessible to the public to conduct experiments on and 2. the inaccessibility of well - researched methods and techniques released.

1.2 Problem Statement

In recent years, topics such as fraud detection and prevention have received a lot of attention on the research front, in particular from payment card issuers. The reason for this increase in research activity can be attributed to the huge annual financial losses incurred by card issuers due to fraudulent use of their card products. A successful strategy for dealing with fraud can quite literally mean billions of dollars in savings per year on operational costs. Fraud prevention is very important to financial institutions. The advent of new technologies such as Telephone, Automated Teller Machines (ATMs) and credit card systems have amplified the amount of fraud loss to many financial institutions. Analyzing every transaction as legitimate or not is very expensive. Moreover, it is also time consuming, hence practically impossible. Confirming whether a transaction was done by client or fraudster is a better option, but phoning all cardholders is cost prohibitive if it

is checked in all transactions and will lead to customer dissatisfaction. In recent times, fraudsters have also been adapting new operational practices to escape early detection due to an increase in credit card fraud detection techniques. Credit card fraud detection models therefore require regular upgrades. This study aims to help issuers of credit cards to reduce credit card fraud to the lowest minimum. This will then eliminate the tremendous losses that these financial institutions suffer annually.

1.3 The Study Objectives

1.3.1 General Objective

The general objective of this study is to identify the most suitable classification algorithm for classifying credit card transactions as genuine or fraudulent.

1.3.2 Specific Objectives

The specific objectives of this study are

- To evaluate the proposed techniques using various input and output parameters such as classification errors, accuracy and false alarms.
- To determine which of these techniques will yield results with the lowest error rate.
- To determine which of these techniques will reduce the mis - classification and the generation of false alarms.

1.4 Methodology

A promising approach to identify credit card fraud is to examine the customer's transaction model. If new transactions deviate from this pattern, such transactions are considered fraudulent. The function that illustrate the customer's behavior is selected from the data used in the study. Transaction number, date, period, location, transaction frequency and billing address are the behavioral characteristics showing the expenditure patterns of the credit cardholders. The models are trained based on these datasets. Support vector machine, Artificial neural networks and the K - nearest neighbor are the machine learning algorithms used in this study for analyzing the data because they are better classifiers compared to other algorithms out there.

1.5 Data Acquisition

For the identification of fraudulent transactions, credit card fraud detection systems are enabled on the basis of the amount of transactions made by card holders. In the form of kaggle datasets, the purchases made by credit card holders are derived. Kaggle datasets are nothing but datasets that are already in existence and are being posted for data mining and machine learning purposes by credit card companies and other researchers. Data was obtained from `http:/www.kaggle.com` database which included data rows of $6,000$ and columns of 31. The only columns that were important for the study are Time, Amount and Class (genuine or fraudulent) out of the 31 columns. The other 28 columns were transformed using principal component dimensionality reduction in order to protect user identities. Kag-

gles are online communities that enable researchers to find and use datasets that have been published for research purposes. Datasets used in credit card fraud identification systems include only credit card holders' transactions.

1.6 The Design of the study

The numerous steps running from the raw data process through to the validation stage are seen from figure 1.1. The process involves the stage of data conversion, stage of data selection, stage of feature extraction, stage of data preparation, cross validation using grid search stage, collection of correct parameters stage, stage of support vector machine classifier, stage of expected outcomes and stage of validation. The following are the explanations of the various stages of figure 1.1

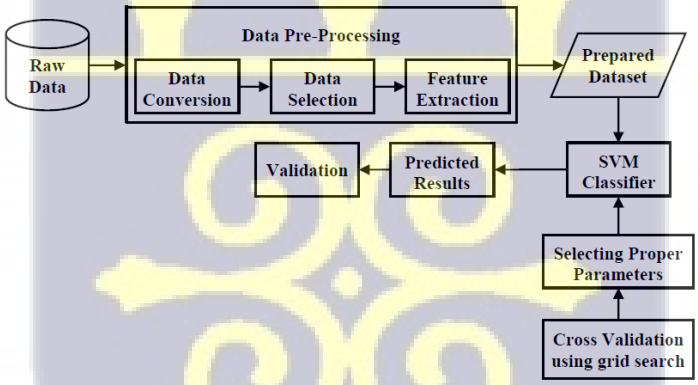


Plate 1.1: Workflow of the Training Model

above which is the flow chat of the design of the study.

- Raw Data stage is the stage at which the raw dataset is fed into the algorithm.

- The data conversion stage is where the features used in the dataset are converted into numerical data.
- The data selection stage is where the algorithm selects the important features in the data to efficiently portray the usage of the behavior of an individual credit card account.
- The feature extraction is a method used to reduce the large size of data ie, a dimensional reduction method. This is where the relevant characteristics of the input dataset are extracted due to the high dimensionality of the dataset.
- The prepared dataset stage is where the huge amount of dataset is processed. This is where a scalable machine learning system is needed to process the large amount of data.
- The SVM classifier is the stage where the data is trained.
- The cross validation using grid search is the stage at which proper parameters are chosen to avoid over fitting of the algorithm to get a perfect result.
- Selecting proper parameter stage is the stage where the parameters C and γ are selected. C is regularization parameter which weighs the classification errors. It is the trade off between genuine transactions and fraudulent transactions. The radial basis function has a parameter known as Gaussian with γ which controls the width of the RBF kernel function. These two parameters are most important for the accuracy of the classification. Effectively selecting the proper values of C and γ avoids over fitting and gives a high classification accuracy.

- The predictive result stage is where supervised learning models are produced by using the knowledge of already classified data to inductively find a predictive pattern of the algorithm.
- The validation stage which is the final stage is the stage where the optimal number of support vectors are determined or a stopping point for the algorithm.

1.7 Computational Tool

MATLAB is the analytical method used for training and identification and statistical evaluation of the fraud detection algorithms. MATLAB is a high - level language, intentionally made for numerical computations. For solving mathematical problems, it makes available appropriate interactive command line interfaces. It is also used as a batch - oriented data processing language. It offers a broad range of graphics capable of visualizing and manipulating data. MATLAB is usually used for analyzing data but can also be used for writing non - interactive programs via its interactive command line interface. The major advantage of MATLAB is that most of its programs are portable.

1.8 Significance of the Study

Though most of the fraud detection systems show good results in detecting fraudulent transactions, they also lead to the generation of false alarms and mis - classification of the dataset. This significance especially in the domain of credit card fraud detection is to enable credit card companies to minimize their losses and at

the same time, do not want the card holders to feel restricted too often.

1.9 The Study Limitations

While detection of credit card frauds have been in place for a long time, in this very important interest of research, not much work has been done. This shot drop is accounted for by the lack of real - world life data available on which researchers can conduct experiments as credit card issuers are not prepared due to security concerns to provide data on sensitive consumer transactions. They even used to modify their interest names so that the investigator could not understand the ideas concerning real interests. In addition, it is very difficult to manage the large quantities of data effectively due to the very high dimensions associated with the available data. The mis - classification costs for credit card fraud detection systems are typically high.

1.10 Research Contributions

Only a few credit card fraud detection models were introduced in the research front as defined in the literature review due to the lack of real life data. And the few models were not implemented into the real systems for detecting frauds. However, some major uses of the few data mining techniques can still be found. Outlier detection, self - organizing maps, neural networks, Bayesian classifiers, support vector machines, artificial immune systems, fuzzy systems, genetic algorithms used in detecting credit card frauds, K - nearest neighbor and hidden Markov chains are some of the data mining techniques that have been used over the years.

The findings are relevant and important as they bring to light the potential benefits there are in using some of the Machine Learning techniques in fraud detection methods. These would be useful for insurance and financial companies as well as credit card and policy analysts who might directly benefit from the results of this research. The tools used in this research can be widely used for monitoring product performance on the markets, track progress of fraud detection tools and countless other applications.

1.11 Thesis Structure

This research was organized into five different chapters. Chapter one provides the study history, problem statement, aims, brief explanation of the methods used, importance of the study, limitations and contributions to research. Certain related literature is discussed in Chapter two. The research methodology is discussed in chapter three. In this chapter, all the methods and mathematical foundations used in the analysis are discussed. The theoretical outcomes, their respective debates and explanations are discussed in Chapter four. All the findings are summarized in Chapter five and conclusions and recommendations made.

Chapter Two

Review of Literature

2.1 Review of Literature

Credit cards fraud detection are crucial parts of the measures put in place for implementing and to maintain attacks tolerant database system. Although data management systems are largely able to serve as intrusion prevention to some level, they are not adequate to defend against syntactically correct yet semantically damaging transactions through conventional access control mechanisms. The detection of credit card frauds have contributed to a lot of interests and a range of approaches with a particular focus on data mining and neural network.

2.1.1 The P - RCE Neural Network

Using credit card frauds detection systems, Vatsa et al. (2005) calculated the efficiency of neural networks. The P - RCE neural network (Restricted Coulomb Energy) was the neural network used for this research. The authors concluded that 20% to 40% in reducing the overall loss due to fraudulent transactions could

be achieved.

2.1.2 CARD WATCH

Aleskerov et al. (1997) proposed CARD WATCH, a data mining system based on neural network learning modules. These systems trained neural networks with the transaction history of a particular cardholder which was used to process a current transaction history and to detect changes that may occur. The authors assumed that the fraudster will purchase so many items within the shortest possible time and that this habit of the fraudster will ensure that any anomalies in the transactions will be detected. If deviations appear, the suspected transactions were taken for special considerations. There is no certainty however that with this procedure every fraudulent transaction can be detected since a cardholder can decide to purchase in a limited time that will deviate from his/her usual transaction history. But because a thief's usual activity is to buy as much as possible in a short period of time, the purchase anomaly would most likely be observed.

2.1.3 Parallel Data Mining Approach

A large dataset of credit card transactions was split into smaller groups by Delamaire et al. (2009) and mining methods were applied in parallel to distributed data mining approaches. To produce a meta - classifier, the outcomes of these individual models were then placed together. A large dataset of credit card transactions was split into smaller groups by Chan et al, and mining methods were applied in parallel to distributed data mining approaches. To produce a meta - classifier, the outcomes of these individual models were then placed together. Credit card frauds

detection models focused on the cardholder's transaction history and used for detecting credit card fraud in their current transactions was proposed by Zhang et al. (2009). To develop the model, the transaction history of individual credit cards were used. The use of unsupervised self - organizing map methods were required by this models to identify outliers from their normal transactions history.

2.1.4 Data Mining Model

A model based on data mining was described by Sun et al. (2011) where web services were used to move data from one bank to another. For credit card fraud detection, algorithms for fraud patterns have been created. This approach allowed banks to share their experiences of fraud trends in both heterogeneous and dispersed environments and to further strengthen their ability to detect fraud and thereby minimize financial losses for these banks. The use of this technique of data mining is well illustrated by Chiu and Tsai (2007). They considered data exchange web services among financial institutions. For mining fraudulent association rules that provide information on the characteristics that occur in fraudulent activities, Fraud Pattern Mining (FPM) algorithms were advanced. By using the latest fraud trends to stop fraudulent attacks, banks improved their original fraud detection systems. Although methods of data mining are generally effective, they are inherently very slow.

2.1.5 Outlier Mining Technique

Yu and Wang (2009) discovered outlier mining techniques for credit card fraud detection. Distance - based outliers were identified and algorithms developed

for outlier mining. By calculating distances and setting thresholds for outliers, the model was able to identify outlier sets. The model has successfully defined overdrafts and has also been used to forecast fraudulent credit card transactions.

2.1.6 Technique of Mining Neural Data

Yu and Wang (2009) made use of the technique of mining neural data. This process focused on the transaction habits of customers. In developing the model, variations from the normal behavior patterns were taken as significant tasks. With the data and confidence value estimated, the neural network was learned. Low - confidence credit card transactions were rejected by the qualified neural networks and were considered as fraudulent transactions. With additional confirmation and the detection performance compared to the thresholds, abnormal confidence values were tested.

2.1.7 Method of Fusion

Panigrahi et al. (2009) proposed a method of fusion. The method consisted of four separate parts: rule - based filter, Dempster - Shafer Adder, database of transactions history, and Bayesian learner. For finding the suspicion rate of the transactions, the rule - based filter was used. In calculating the initial belief, based on the evidence presented by the rule - based filter, the Dempster - Shafer theory was used. Depending on the initial assumption, the transactions were classed as normal, abnormal or suspicious. Once a transaction was found to be suspicious, according to its similarities with the transaction history of the cardholder using the Bayesian learning, belief was further enhanced or weakened.

2.1.8 Stochastic Methods

Extensively simulating stochastic methods show that, relative to other models, fusion of different evidence have strong positive effect on the performance of credit card frauds detection systems. There are some technological challenges to the production processes of fraud detection systems created by related data characteristics, such as large data volumes, distorted distributions, irregular transaction costs and strongly overlapping groups. Building training data for the classifiers is the most severe one. Particularly, because of unstable learning algorithms such as neural network and decision tree, for the classifiers. In response to minor changes in training data, the learning outcome is subject to substantial changes. For these purposes, in the first stage of modeling fraudulent transactions, the generation of unbiased data for the training of classifiers is very significant. However, due to the distorted nature of data delivery as well as their large sizes, it is very difficult. The concept is essentially to enhance the efficiency of neural classifiers that are constructed with training data consisting of only minute sections of actual transaction data and are highly biased towards fraudulent transactions. Since bias of data training is expected, the approach is constructed with frauds density maps using analyzed data over real data in the design of the neural classifiers for detecting fraudulent transactions. From the understanding that input transactions can be considered suspicious if the likelihood of fraud occurrences is represented by fraud - ridden areas in the input space and the frauds density, then the fraud ratio of the neighborhood surrounding the points represented by an input vector in the input data space is taken as a confidence value with the neural classifiers' fraud score modified.

2.1.9 Imbalance Data Classification

Credit card transactions were considered by Panigrahi et al. (2009) as an imbalanced data classification issue where most samples (non - fraud) are much more than minority samples (fraud). The algorithms perform very poorly when handling imbalance data, and the results are skewed to the majority class of the data. Therefore, for the classification of imbalanced data, an appropriate method is needed, particularly for credit card fraud detection problems. For this purpose, another model for detecting credit card frauds was developed that used hybrid sampling techniques to over - sample the data, consisting of random under - sampling and Synthetic Minority Over - Sampling Techniques (SMOTE).

2.1.10 Credit Card Fraud Identification

Pimentel et al. (2014) recently did an impressive report on credit card fraud identification. The study revealed that the key types and sub - types of identified frauds were formalized by trained fraudsters and the existence of data evidence gathered within affected industries was presented. In order to create a more precise global method, Chan et al. (1999) outlined a meta - classifier framework for the detection of credit card frauds by combining the results obtained from local fraud detection instruments at various corporate sites. A comparable study was performed by Chan et al. (1999) and established by Chan et al. (1999). To complement the various classification findings, this work described more practical cost models. The combination of several classification rules was also suggested by Wheeler and Aitken (2000). A fraud detection technique was introduced by Pimentel et al. (2014) using stacking bagging approaches to boost cost savings.

There are numerous forms in which fraudsters demonstrate interest in showing that fixed credit card numbers are valid and can be used for fraudulent activities. This is due to the poor existence of conventional credit card processing systems, where long - term, semi - secret and often fully disclosed keys used for authentication can be used. While credit cards are used by a large number of individuals worldwide, there are no alternative methods of payment that are more reliable. By using the notion of disposable credit card numbers, this issue was resolved. In that event, a certain number to be used for transaction purposes was decided between the credit card issuer and the holder and this number was discarded just after the transaction and a new number was created for the next transaction and so on. The fraudsters are therefore unable to sniff an identification number that would help them collect any useful data.

2.1.11 Artificial Neural Networks

Hussain et al. (2008) have introduced an Artificial Neural Networks (ANNs) based fraud engine. It has been applied to credit card purchases in order to gain insight to their efficiency and the general viability of their approach. The explanation is that security measures focused on intelligence behavior may provide additional protection for sensitive systems. Real - time identification and response to fraudulent activities is of particular interest. Any suspicious or irregular operations are quickly captured and avoided. For actual artificial implementation using Neural Networks, protection measures based on actions have shown that they are at the same level as the attacker and not a step behind. Using this type of technique as protection takes it down to the personal level, according to the transaction history,

meaning frauds should be observable for each single user or client.

2.1.12 Neural Fraud Management Systems

There are existing systems that are designed for examining the transaction history of cardholders for the identification of potential frauds. These existing systems are referred to as fraud engines and are focused on analyzing the history of transactions to determine if a new history is found. Highly automated Neural Fraud Management Systems (NFMS) are state - of - the - art integrated neural network systems, fraud detection engines and automatic modeling systems. In order to resolve the limitations of ruled - based systems, machine learning, anomaly - detection, Artificial Neural Networks systems are typically used. These systems are able to automatically and instantly detect suspicious activity for each cardholder and for groups of cardholders, rather than having to wait for a new attack to be identified and for a new rule to be written by experts. Behavioral systems are inherently future proof, since the first time they are launched, they can spot new kinds of attacks. Clustering algorithms based on artificial neural networks depend on successful anomaly detection systems. Clustering algorithms group a smaller number of clusters into identical transactions. A certain behavior pattern is denoted by each cluster. Each time the anomaly detection system processes a new transaction, the systems try to fit it into an existing cluster. It is listed as fraudulent if a transaction does not fit into any of the clusters.

2.1.13 Optimal Margin Classifier

Cortes and Vapnik (1995) presented an optimal margin classifier training algorithm in which he showed that increasing the margin between the training examples and the class boundary resulted in reducing the maximum loss of these classifiers in terms of generalization efficiency. His definition was originally explored because binary class optimal margin classifiers achieve less separation of the training data error, since separation is possible, and outliers in such classifiers are easily detected. Separable data were the basis of the first inquiries into these types of algorithms. But these algorithms were expanded by Cortes and Vapnik (1995) to account for linearly inseparable datasets. Attempts were made to apply these outcomes to issues of multi - class classification. This type of learning machine was later named the Support Vector Machine (SVM). A machine learning technique with a clear and sound theoretical context is the support vector machine. It is important to note that researchers have reported that support vector machine in binary classification problems equal or outperform neural networks in most cases. Simulated annealing, multi - grid random search, time - weighted pseudo - Newton optimization, and discrete propagation of error were gradient - based methods. In their experimental problems, simulated annealing outperformed all the other techniques. These approaches need a much more training time than the others, with an increase in their training time as sequence lengths increase. Most attempts to close the gap between long time lags can be outperformed by simple random weight guessing, as stated earlier. Some techniques have promised good results for longer time lags, such as unit addition and Kalman Filter ANNs, but these are not useful to problems of undefined lag length. If properly imple-

mented, information technology can help identify and handle credit card frauds in very powerful and efficient ways. Several techniques and technologies have been proposed for the implementation of Fraud Management Systems (FMS), the most popular of which are rule - base scoring models, computational methods, and Artificial Neural Networks (ANNs). Although artificial neural networks have some advantages and are commonly used for the detection of credit card fraud, conventional back - propagation - based algorithms used for training artificial neural networks have local optimal problems that decrease the effectiveness of artificial neural networks in detecting credit card fraud. Scoring models are a set of rules defining the features of certain forms of documented fraud. Scoring models work well when the kinds of fraud in the size and complexity of rule parameters are clear and manageable. In the case of simple forms of fraud, the description of all the laws involved can be rather boring. Scoring models often have no self - learning skills and do not readily implement new ways of fraud. Several statistical approaches, such as the Least - squares regression analysis, are effective in detecting credit card fraud. In this situation, auditors who are only interested in fraud detection systems without paying attention to the mathematical specifics involved need some guidelines. In the development and training of back - propagation neural network models, hybrid financial models such as trend analysis models have been proposed. These studies have shown that these suggested models provide a very high rate of prevention and often outperform other models of fraud detection, such as discriminant analysis , decision trees, and neural networks of back - propagation. Furthermore, in the management of fraud schemes, complete statistical approaches do not easily provide adaptive learning capabilities. Artificial Neural Networks (ANNs) with self - learning skills are suggested to solve the above -

mentioned problems. Artificial Neural Networks learn and can make decisions by experience extended to new ones from previous experiences. They are inspired and work like neurons in the human brain by information - processing units. Neural networks are known to be non - parametric black box classifiers consisting of artificial neurons. They are flexible to use. Therefore, Artificial Neural Networks have become useful developments in fraud detection systems for pattern recognition and implementation. The performance of artificial neural networks, considering their increasing distribution, depends largely on how these networks are equipped. Training is designed on the basis of Back - Propagation (BP) algorithms for most artificial neural networks for fraud detection. These algorithms suffer from well - known problems, so local optima can be difficult to solve. This is because the starting point in an n - dimensional space is initialized randomly by all BP algorithms used for training artificial neural networks. A BP algorithm possibly converges on a local solution if the starting point is located in a local valley. Several reforms have been made to solve this issue by seeking a global solution. In recent times, by optimizing artificial neural network environments, certain fraud detection systems use ant colony algorithms. They function using the steps below:

1. Usage of personal identification numbers (PINs) and passwords for initial authentication.
2. Detecting automatic fraud using artificial neural networks
3. Automatic fraud detection using artificial neural networks

Stage 1 is made up of preliminary authentication processes for verifying user identities. Unique Identification Numbers (PINs), passwords, or Internet Procol

Addresses, fingerprints on some terminals, and telecommunications identity sim cards can be used. Phase 1 will bring about a large number of problems with identity misuse (say identity theft). The most significant portion of this procedure is Step 2. Step 1 is a method of signature - based analysis, step 2 is a behavior - based process that needs more smart approaches in the analysis of the dataset on behavior. In order to design their trends, data on the transaction history of the card users is collected and analyzed. The case for manual analysis in phase 3 is then illustrated for any transaction that deviates from the transaction history. Abnormal transactions, however, are just possible illegal events and are sometimes just rarely used by legal consumers. Thus, in phase 3, the suspicious transactions observed in step 2 are set up for further analysis by the appropriate workers. This three - step approach thus forms a framework strategy that provides the individual in charge of fraud detection with overall controls.

2.1.14 Artificial Immune Systems

Artificial Immune Systems (AIS) are very effective biological systems - motivated strategies. In many fields, such as computer protection, optimization, robotics, fault detection, etc; the applications of artificial immune system models are increasing significantly in different domains. But only a few attempts have been made in this field to identify credit card fraud. Gadi et al. (2008) indicated that when the information is extremely distorted, Artificial Immune System algorithms outperform some conventional methods of detecting credit card frauds. Brabazon et al. (2010) have established that it is possible to use artificial immune techniques to prevent credit card fraud online. In this work, they examined the efficacy of AIS

in the detection of credit card frauds. The findings indicated the potential of AIS algorithms to be used in fraud detection systems.

2.1.15 Questionnaire - Responded Transaction

Instead of using credit cardholders' transaction history to detect credit card fraud, Chan et al. (1999) introduced a new approach for solving credit cards fraud issues. To detect credit card fraud, they built customized models. Cardholders' personal transaction information was first gathered using an online self - completion questionnaire system. On line framework Questionnaire - Responded Transaction (QRT) datasets were considered as transaction records and used to create customized models that are used sequentially to detect whether or not a new transaction is actually fraudulent. Since the spending conduct of the unauthorized user is often different from that of the cardholder, fraud can be prevented even without any transaction dataset from the initial use of credit cards. The method of the QRT is very encouraging. There are however, several issues associated with it that need to be addressed. One of the most important questions about the use of the QRT method is how to detect accurately with just a few datasets, say 150 to 300, because card users typically do not want to answer too many questions.

2.1.16 Fraud Detection Systems

Fraud Detection Systems (FDS) have recently been categorized into two major categories: abuse - based and anomaly - based. New fraud patterns are not observable by the misuse - based FDS. Anomaly - based detection systems, however, often raise several false alarms. It uses a series of aligned strategies to solve prob-

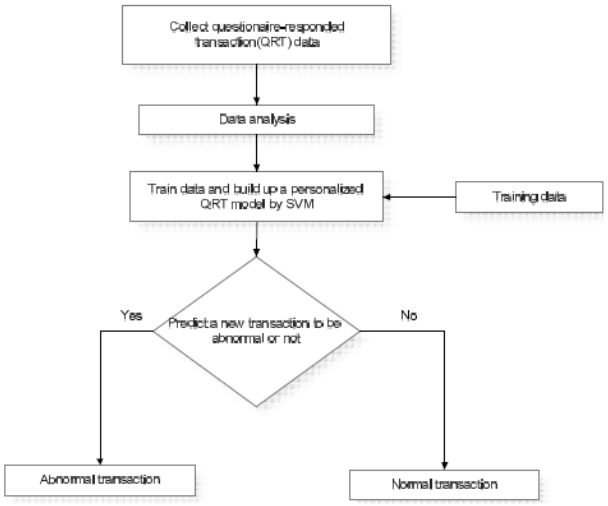


Plate 2.1: QRT Approach for Predicting Credit Card Frauds

lems associated with credit card fraud in evaluating potential cases of fraudulent transactions. The strength of both misuse and anomaly - based detection techniques defines this type of model. This function of the FDS makes it possible to address the shortcomings of current fraud detection systems and makes it effective in the detection of credit card fraud. Sequence alignment was used to evaluate the similarities of both the genuine sequence of cardholders and the sequence created by a legitimate fraud detection model of an incoming sequence of transactions. In deciding whether or not a transaction is fraudulent, scores from these two phases are combined. Knowing that behavioral models and classifications at the transaction levels are not simple strategies for detecting credit card frauds, it is helpful to find other approaches that can improve their strengths. Along with other current rule - based detection systems, all the techniques can apparently be put to use. In a given period of time, an extension to transaction - level classification is to collect information on various transactions. This is mainly accomplished by designers

of detection systems for commercial fraud. Previously, supervised classification - based approaches to credit card fraud transactions were published with a focus on classifying individual transactions rather than account level transactions.

2.1.17 Secret Markov Model

Srivastava et al. (2008) for credit card fraud detection, came up with a Secret Markov model. In this model, a Hidden Markov Model (HMM) was used to model the sequence of operations in credit cards processing. An HMM was initially trained by the cardholder's regular actions. If the qualified HMM does not accept an incoming transaction with a sufficiently high probability, then this incoming transaction is assumed fraudulent. This model seeks to ensure that real transactions are not rejected.

2.1.18 The Ada Cost

For credit card fraud detection systems, Lee et al. (2001) suggested the application of distributed data mining. This approach used Ada Cost, which uses a large number of classifiers and required a lot more computational resources during detection, to increase the effectiveness of highly distributed databases and detection systems.

2.1.19 Naive Bayesian Approach

In addition to a low false alarm, Brause et al. (1999) combined advanced data mining techniques and neural network algorithms to achieve high fraud detection. For credit cards fraud detection, the Naive Bayesian approach has also been proposed.

In many real world datasets, the Naive Bayesian algorithm is very effective and linear attributes are extremely efficient. Bayesian networks were more precise and quicker to train, but when applied to new situations, they are slower.

2.1.20 Distance Amount Outlier Detection

Chaudhary et al. (2012) proposed a study based on distance amount outlier detection mining on the credit card fraud detection model, which shows that outlier mining can detect credit card fraud better than clustering - based anomaly detection. Outlier mining was used in credit card fraud detection in this model. Definitions of outliers based on distance are referred to and the outlier algorithms have been developed. By calculating distances and setting thresholds for outliers, this model detects outlier sets.

2.2 Experimental process

1. Input was standardized with the standard method of deviation
2. Distances are calculated between two datasets of credit card transactions and the total value is collected. It is called an outlier if the sample value is higher than other objects.
3. Threshold setting of the outliers. It is then compared to the sets obtained from the outlier. This technique can provide precise forecasts, but when applied to large datasets, scalability and efficiency problems occur.

The available methods of detecting credit card frauds using labeled data in training their classifiers do not identify new forms of fraudulent transactions. The most

popular drawback of supervised methods of learning is that they require human intervention in parameter optimization. Decision trees, however, do not need the user's parameter settings and can be created quicker than other methods.

2.2.1 The BOAT Algorithm

The BOAT algorithm is an effective fraud detection system which through the combination of clustering and classification methods, adapts to changes in behavior. To detect any irregularities at the first stage, it is a two stage fraud detection method for comparing an incoming transaction to the transaction history. The next step is to reduce the incidence of false alarms. Frauds history databases tested suspicious anomalies and guaranteed that these anomalies found were due to fraudulent transactions or some short - term changes in the spending pattern. BOAT assisted gradual updating of the transaction database in that work and managed full coverage of fraud with high speed and less expense. On both synthetically generated and real - life data, the suggested model is assessed and has shown very good accuracy for the detection of fraudulent credit card transactions. As discussed above, the majority of fraud detection systems show a number of variations in their precision outcomes. The key problem that most of these systems have found is that the majority of transactions flagged by these fraud detection systems as fraudulent are simply not fraudulent. Banks expend a lot of time and money on investigating a great deal of legitimate cases. Customer inconveniences and disappointment are often induced.

2.2.2 Intrusion Detection System

Axelsson (2000) clarified that the factor limiting the efficiency of intrusion detection systems is not the ability to correctly recognize the invasive actions, but its strength to reduce false alarms, because of the base - rate fallacy problems. Although failure to identify a fraudulent transaction results in the company's direct loss, follow - up measures taken to chase false alarms often appear to be expensive. Any design selection that aims to increase the rate of accurate fraud detection typically false alarms often trigger an increase. One of the reasons of this research is to answer this problem. It is well established that each holder of a credit card has a certain shopping pattern that creates for him an activity profile. In subsequent transactions, almost all current fraud detection systems aim to catch these activity patterns as rules and search for any breaches. These laws are, however, essentially static in nature. As a consequence when the cardholder discovers new activity patterns that the fraud detection system is not yet aware of, they become ineffective. The aim of efficient fraud detection systems is to dynamically train the actions of users to reduce their own losses. Systems that can not adapt or 'learn' can quickly become redundant, resulting in a large number of false alarms. Fraudsters can also try new forms of attacks that the fraud detection system should still detect. For example, by making a few high value transactions or a large number of low value transactions so as not to be detected, they can seek to derive maximum benefits. Fraud detection systems should therefore be established that can incorporate various facts, including patterns of legitimate cardholders as well as fraudsters. Credit cards fraud detection statistical techniques were reviewed by (Bolton and Hand, 2002).

2.2.3 Supervised and Unsupervised Techniques

Techniques for detecting statistical frauds are divided into two major groups: supervised and unsupervised. In the case of controlled fraud detection systems, when classifying new transactions as legitimate or fraudulent transactions, models are determined based on the transaction history. Outliers or transactions which deviate from the cardholder's transaction history are considered as potential cases of fraud in unsupervised fraud detection systems. To estimate the chances of fraud in a given transaction, all fraud detection techniques are used. Models of credit cards fraud detection are practically in active use today. Taking into account the proliferation of data mining methods and applications in recent years, only fewer works in the interest of credit cards fraud detection have been published. One paper recently considered many fraud detection strategies in credit cards fraud detection, such as support vector machine and random forest. Their work concentrated on the effect of the aggregation of transaction data on fraud detection results. The research analyzed the aggregations on two distinct real - life datasets over time differences and discovered that aggregation can be beneficial with critical factors being aggregation time lengths. With random forests, aggregation was very successful. Random forests revealed findings that were higher than other approaches used. Logistic regression and support vector machine also performed well. Complex data mining techniques that demonstrate superior performance in distinct applications are support vector machine and random forest.

2.2.4 Support Vector Machine

Support vector machine is a tool used in solving a variety of problems for statistical learning purposes that have very effective histories and are useful applications. They are equivalent to neural networks. They are used as alternatives for obtaining neural network classifiers by the use of kernel tricks. The support vector machines aim to reduce the upper bound on the generalization errors instead of reducing errors that can be seen in the data. Support vector machines can be used to achieve general solutions with good globalization errors when contrasting support vector machines to other statistical approaches such as neural networks which suffer from local minimum, over - fitting and noise. In applications that have in - built model choices in the optimization process and have been found to perform better than neural networks in solving classification issues, support vector machines are more convenient. To be able to achieve good results in the application of support vector machines, the required selection of parameters is important. Problems related to some of the techniques mentioned in the literature are that a labeled data is a prerequisite for them to work effectively in classifier training. The greatest problem connected with credit card fraud detection techniques is collecting real - life data. Often, new fraudulent transactions in which labeled data are not used in the study can not be identified by these methods.

2.3 Gaps In Literature

I have discussed some credit card fraud detection algorithms. The gaps in the implementation of these algorithms for credit card fraud detection have also been

revealed with possible modifications to overcome their setbacks. All the algorithms seek to reduce credit card frauds. It is important to note that most of these credit card fraud detection techniques could not meet their usefulness at the time of their use because they were being test run and also due to the non - availability of credit card data.

2.4 Choice Of Algorithms For This Study

The choice of SVM, ANN and the KNN algorithms for this study over the numerous algorithms out there for credit card fraud detection is that they are more robust and produce more accurate results compared to the numerous algorithms out there due to technological advancement.

Chapter Three

Research Methodology

3.1 Introduction

A thorough analysis of Support Vector Machine(SVM), Artificial Neural Networks(ANNs) and K - Nearest Neighbor(KNN) are presented in this chapter. Support Vector Machines are methods used for classification purposes in pattern recognition. Classifying trends into two types is a classifier; fraudulent and non - fraudulent transactions. For binary classifications, it is quite suitable. Support Vector Machines are model - free techniques that provide successful solutions without any assumptions about the distribution and inter - dependence of the data for classification problems. They are intelligent artificial tools that have to be conditioned to acquire learned models. In 1995 Cortes and Vapnik introduced Support Vector Machine as a statistical machine learning method as an alternative to polynomial, radial functions and multi - layer perception classifiers, in which the weights of the neurons are identified by solving Quadratic Programming (QP) problems with constraints of linearity and equality rather than solving non - convex, unconstrained minimization problems. As a new machine learning method,

regression analysis, face detection, categorization of texts, data mining and detection of outliers for binary classification purposes, support vector machine faces challenges if the data are very large due to the dense design and memory requirements of the quadratic datasets. However, support vector machines are excellent examples of supervised learning that seek to optimize generalization by maximizing the margin and use kernel functions to facilitate nonlinear separations. Support vector machines aim to prevent over - fitting and under - fitting from occurring. Margins in the support vectors denote the distance in the feature space between the boundary and the nearest data points.

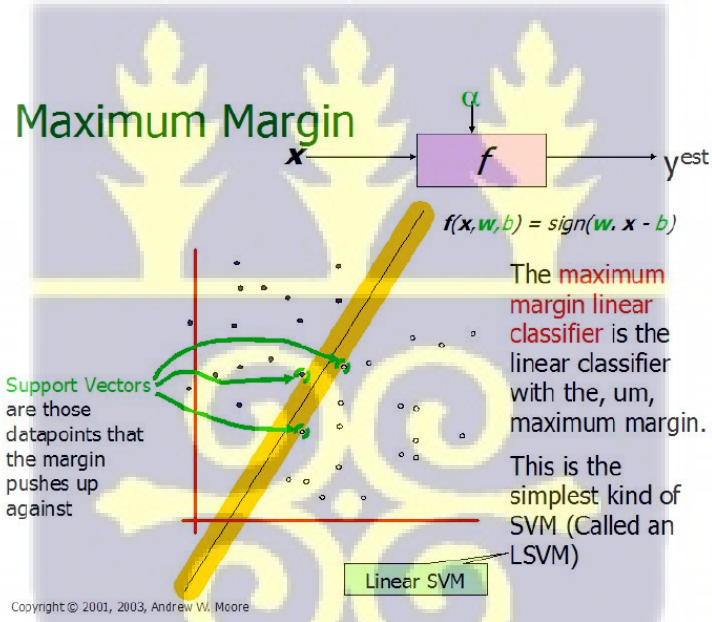


Plate 3.1: An illustration of a Linear SVM (Taken from Andrew W. Moore slides 2003)

The mathematical background of support vector machines is described in details as follows: They were built from the theory of Structural Risk Minimization problems. For binary classification problems, the decision function of a typical support

vector machine is as shown in Equation (3.1)

$$f(y) = sgn(\omega^T y) + c \tag{3.1}$$

in which y is an example to be classified, ω is the input vector with weights and c a bias. Equation (3.1) is used in the determination of a decision boundary between the two classes. The parameters ω and c have to be learned by the support vector machine on the training phase and c is derived by maximizing the margin of separation between the two classes. The criteria used by SVM is based on the margin maximization between the two classes. The margin is that distance between the two hyper - planes. To find the hyper - plane,

$$A : z = \omega^T y + c = 0 \tag{3.2}$$

and the two hyper-planes

$$A_1 \colon z = \omega^T + c = 1 \tag{3.3}$$

and

$$A_2 \colon z = \omega^T y + 1 = -1 \tag{3.4}$$

The starting point separating the two classes is A and the two margin boundaries are A_1 and A_2. Then the margin is

$$\frac{2}{||w||}$$

, where $||w||$ is the norm of the vector ω. In a non perfectly separable case, the margin is soft. There is therefore the tendency of introducing mis - classification errors. This mis - classification errors when noticed, should be minimized. They are minimized by introducing a slack variable ξ_i. If ξ_i=0, the classes are correctly separated. Where ξ_i is a non - negative slack variable for mis - classification. Let z be the indicator of the class, where in the case of fraud detection $z = 1$ for the positive class and $z = -1$ for the negative class. SVMs require that either

$$(\omega^T y + c) \geq 1 - \xi_i$$

or

$$(\omega^T y + c) \geq -1 + \xi_i\}$$

which is summarized as in equation (3.5)

$$z_i(\omega^T y + c)\& \geq 1 - \xi_i \tag{3.5}$$

where $i = 1, 2, \cdots, n$

The optimization problem for the calculation of w and c can thus be defined by

$$Min \frac{||\omega||^2}{2} + C \sum_{1}^{n} \xi_i$$

subject to

$$z_1(\omega^T y + c) \geq 1 - \xi_i$$

with $\xi_i \geq 0$

By minimizing $\frac{||\omega||^2}{2}$ the complex support vector machine is reduced and by minimizing the slack variable, the mis - classification errors are reduced as well. C is a regularization parameter that weighs the classification errors and is the trade - off between the two classes. The constrained optimization problem is solved by using the Lagrange function as shown in equation (3.7).

$$L(\omega, c, \xi, \alpha, \beta) = \frac{||\omega||^2}{2} + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\big(z[\omega^T y + c] - 1 + \xi_i\big) - \sum_{i=1}^{n}\beta_i\xi_i \tag{3.6}$$

The solution of the optimization problem is given as minimizing w, c, ξ and maximizing α and β. It is easier to solve the problem by introducing it's dual formulation in equation (3.8),

$$\max_{\alpha\beta}\omega(\alpha, \beta) = \max_{\alpha\beta}\left\{\min_{\omega,c,\xi}(\omega, c, \xi, \alpha, \beta)\right\} \tag{3.7}$$

By substitution, the problem is changed into a dual formulation signified by;

$$max\left\{\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j z_i z_j\langle y_i, z_j\rangle\right\} \tag{3.8}$$

and is maximized under the constraints, $\sum_{i=1}^{n}\alpha_i z_i = 0$ and $0 \leq \alpha_j \leq C$ for $i = 1, 2, \ldots, n$

The Khun -Tucker condition in Equation(3.8) is introduced in Equation (3.9)

$$\alpha_i\left\{z_i(\omega^T y_i + c) - 1 + \xi\right\} = 0 \tag{3.9}$$

where $i = 1, 2, \cdots, n$ The Lagrange vectors are the support vectors required to describe the hyper - planes. In a linearly separable data, all the support vectors lie

on the same margin. The decision boundary is determined by Equation (3.10).

$$f(y) = \sum_{i=1}^{Ns} \alpha_i z_i(y, y_i) + c \tag{3.10}$$

where y is the input vector, (y, y_i) is the inner dot product. Ns is the number of support vectors, and c a bias term. For nonlinear datasets, kernel tricks are used to map the input vector to higher dimensional vector spaces. Kernel tricks are used in other applications as they provide a simple bridge from linear to non - linear algorithms which can be expressed in dot products. Functions that satisfy Mercer's conditions by Vapnik are used as kernel functions. Kernel functions are introduced as

$$(y_i, y_j) \longrightarrow k(y_i, y_j) \tag{3.11}$$

Introducing kernel functions make it possible for the input vectors to be mapped to higher dimensional vector spaces. Some kernel tricks normally used in support vector machines are;

3.1.1 The Linear Kernel Function

$$k(y_i, y_j) = y_i^T y_j \tag{3.12}$$

Linear kernels are the simplest forms of kernel functions. They are given by the inner dot product $\langle$ y,z $\rangle$ and a constant c.

3.1.2 The Sigmoid Kernel Function

$$k(y_i, y_j) = \tanh(\alpha y_i y_j + a) \tag{3.13}$$

where α and a are the parameters used in Sigmoid Kernels. If $\alpha > 0$, then α is acting as a scaling parameter of the input data, and a as a shifting parameter that guides the threshold of mapping. If $\alpha < 0$, the dot product of the input data is not only scaled but also reversed.

3.1.3 The Polynomial Kernel Function

$$k(y_i, y_j) = [(y_i^T y_j) + 1]^d \tag{3.14}$$

Polynomial kernels are non stationary kernel functions. They are suitable for problems with normalized training data.

3.1.4 The Radial Basis Kernel Function

$$k(z_i, z_j) = \exp(-\alpha \|z_i - z_j\|^2) \tag{3.15}$$

The adjusted parameter α plays a very important role in the performance of the kernel functions and should be carefully chosen to meet the requirements of the problem being solved. The above mathematical formulations form the basis for the start of credit cards fraud detection as the decision support tool for detecting and classifying credit card transactions. Recently, decision making activities of knowledge - intensive enterprises depend mainly on the successful grouping of data patterns, despite time and computational resources needed to achieve accurate results due to the complex nature of the datasets and also, their large sizes.

3.2 Variable Selection

Variable selection is a special way of reducing the dimension of the dataset. It is a transformation of the dataset into a reduced form of variables. The variables denote the essential characteristics of the dataset. Aside using the full dataset, we can use a reduced set of it which still contains the important variables of the original dataset. This process is best done using Principal Component Analysis (PCA). PCA is a method of reducing large sized datasets into a much more smaller size without loosing the important variables as in the original dataset. By using PCA in the dataset reduction process, the cost in training and testing in the support vector machine is reduced drastically. PCA applies orthogonal transformations in transforming dataset of possibly correlated variables into a dataset of uncorrelated variables called principal components. PCA is done by using eigen value decomposition of the covariance matrix of the dataset, usually after mean centering the dataset for the variables. The covariance matrix is calculated as shown below

$$Cov(x, y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{3.16}$$

Where x_i and y_i are the values of variables, $\bar{x}$ and $\bar{y}$ are the mean values of x and y, and n is the sample size.

3.2.1 Model Building

For the performance analysis, the SVM training and testing dataset was added to the SVM algorithm. As a single function for the linear, polynomial and radial basis kernel functions, and the inbuilt Matlab codes, SVM classifiers were im-

plemented. The fraud data for classifier preparation, testing and validation was partitioned. 70% for preparation and 30% for research are used for the dataset. With a ten - fold cross validation for each kernel, the linear, polynomial and radial basis function SVM classification kernels were used and the results averaged.

3.3 Artificial Neural Network

Artificial Neural Networks (ANNs), in their simplest forms, are imitations of the human brain. The human brain has the strength to learn new things, to adapt to new conditions that are evolving. The brain is the most impressive capacity to interpret and make its own assessment of incomplete and vague fuzzy knowledge. We may read the handwriting of others, although the way they write might be entirely different from the way others write. Children should recognize that a circle is both the form of a ball and an orange. The infant, a few days old, has the understanding to identify his mother by touch, voice and scent. From a blurry picture, we can recognize a known person. The brain is an organ that is extremely complex and which regulates the entire body. Also the most primitive animal's brain has more power than the most sophisticated machine. The role of the brain is not only to control the physical activities of the body, but also to control things that can not be defined in physical terms, such as thought, visualizing, dreaming, imagining, learning, etc.

3.3.1 Overview of Artificial Neural Network

Artificial neural networks consist of neurons called processing units. Artificial neurons attempt to replicate the structure of natural neurons and their behavior.

Inputs (dendrites), and one output (synapse via axon) form a neuron. The neuron has a role that decides the neuron's activation. $X_1, X_2, \cdots X_n$ are the inputs to the neuron. A bias is also added to the neuron along with the inputs. Usually the bias value is initialized to 1. $A_1, A_2, \cdots, A_n$ are the weights. Weight is a connection to the signal. The product of weight and input gives the strength of the signal. Neurons receive multiple inputs from different sources and have single output.

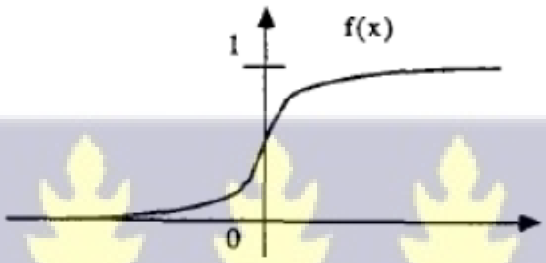


Plate 3.2: Sigmoid function

Various functions are used for activation. The most commonly used activation function is the sigmoid function given by

$$F(x) = \frac{1}{1 + \exp^{-sum}} \tag{3.17}$$

where

$$sum = \sum_{i=0}^{n} X_i A_i \tag{3.18}$$

Phase Functions, Linear Functions, Ramp Functions, Hyperbolic Tangent Functions are other forms of functions used. The function of the hyperbolic tangent ($tanh$) is similar in type to the function of the sigmoid, but its limits are from -1 to 1, unlike the function of the sigmoid, which is from 0 to 1. The sum is the input weighted sum multiplied by the weights between one and the next sheet.

The activation function used is a sigmoid function, which is an approximation of a phase function continuously and differentially. The neural network forms an interconnection between these individual neurons. The artificial neural networks architecture is made up of:

- input layer: for receiving the input values
- hidden layer(s): A set of neurons between input and output layers. The layers consist of single or multiple layers.
- output layers: They have a neuron and their outputs range between 0 and 1.

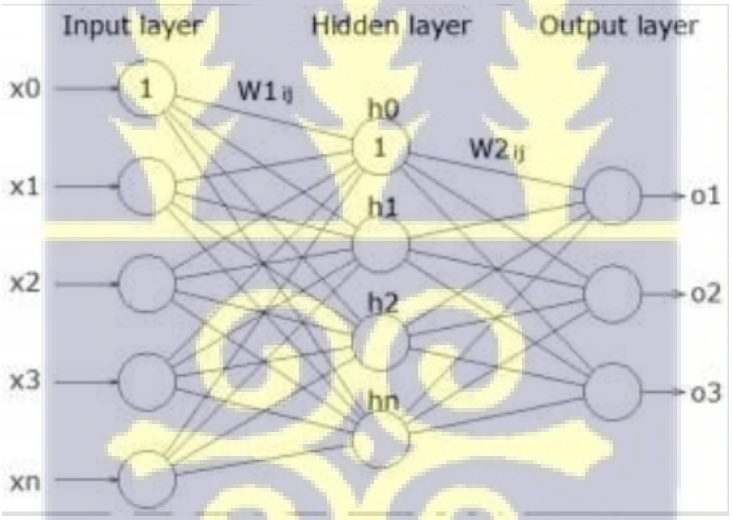


Plate 3.3: Neural network architecture

In inter - unit relation strengths called weights, the processing power is processed. The input intensity depends on the value of the weight. Positive, negative or zero may be the weight value. Negative weight reflects the reduction or inhibition of the signal. Zero weight denotes that the two neurons have no relation. To obtain their appropriate outputs, the weights are modified. To get the necessary

outputs, algorithms are used to change the weights of artificial neural networks. The process of adjusting the weights is known as training or learning.

3.3.2 Model Building

For each function from "$V1$" to "$V28$", pattern recognition networks with 30 input neurons, one neuron was used with one each for "Amount" and "Class". In the hidden layer, 10 to 15 neurons were used and the 10 neuron network provided the best result (Plate 3.4). 70% of the fraud dataset was used as a training set, 15% as a validation set and the remaining 15% for generalization of the network. With Neural networks in Matlab, pattern recognition networks with 30 input neurons, 1 hidden layer with 10 neurons and 1 output layer with 1 neuron were generated and the results averaged. A pattern recognition neural network was formed and the network was trained with the dataset.

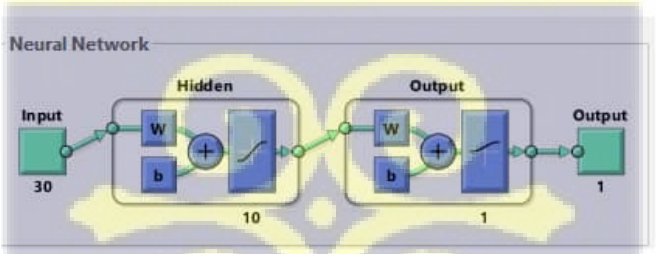


Plate 3.4: Network Architecture

3.4 K - Nearest Neighbor Algorithm

In some anomaly detection methods, the notion of nearest neighbor analysis was used. K - nearest neighbor algorithms, which are supervised learning algorithms, are some of the best classifier algorithms used in the detection of credit cards

frauds. The result of the new instance question is graded based on most of the category of K - Nearest Neighbors.Aha et al. presented it for the first time. The performance of KNN algorithms was influenced by three factors:

- The distance metric used to locate the nearest neighbor.
- The distance rule used to derive a classification from k - nearest neighbor.
- The number of neighbors used to classify the new space.

Among the various credit card fraud detection techniques of supervised statistical pattern recognition, the K - Nearest Neighbor rule achieves consistently high performance without prior assumptions about the distributions from which the training examples are drawn. A distance or similar measure identified between two data instances requires K - nearest neighbor - based credit card frauds detection methods. The KNN is processed by measuring the nearest point to an incoming transaction. Then, if the closest neighbor is fraudulent, a fraudulent transaction is implied by the transaction. The value of K is usually used as a small and odd number to sever the links,(1, 3, and, 5). The effect of noisy datasets can be minimized by large K values. In this algorithm, the distance can be determined in various ways between two data instances. Euclidean distance is a good alternative for continuous attributes. A simple matching coefficient is commonly used as categorical attributes. For multivariate results, each attribute's distance is typically determined and then combined. By optimizing distance metrics, the efficiency of the KNN algorithm can become better. For training, this method requires both legitimate and fraudulent samples of data. Along with high false alarm, it is a fast technique.

3.4.1 Data Preparation

The dataset was retrieved from `http://www.kaggle.com` and is made up of transactions made by credit cardholders in September 2013 by European cardholders. The dataset represents transactions that took place in 60 minutes, where there were 20 frauds out of 6000 transactions. The dataset is highly skewed with the negative class (frauds) accounting for 0.33% of the total transactions. Due to confidentiality issues, it was not possible to get the original features and more background information about the variables in the dataset. The features $V_1, V_2 \cdots, V_{28}$ are the principal components obtained using principal component analysis. The only features of the dataset which have not been transformed using principal component analysis were "Time" and "Amount". The feature "Time" contains the seconds that elapsed between each transaction and the previous transaction in the data. The feature "Amount" is the transaction amount. The feature "Class" is the response variable and takes the value 1 in the case of fraud and 0 if not fraud.

3.5 Evaluation

Accuracy values though important in credit card frauds detection, false alarms and fraud catching rates are the best measurements for credit card frauds detection. In this study, a confusion matrix is used for assessing the fraud catching rates. A confusion matrix is a graphical representation of the True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN). A standard confusion matrix is shown in Plate 3.5 below:

In the confusion matrix, the columns represent the predicted class and the

		Predicted	
		Positive	**Negative**
Actual	Positive	TP	FN
	Negative	FP	TN

Plate 3.5: Confusion Matrix

rows represent the actual class. TP is True Positive(Fraud catching rate) shows the number of genuine transactions correctly identified as non - fraudulent. FP is False Positive (False alarm rate) gives the number of genuine transactions incorrectly identified as fraudulent. FN is False Negative represents mistakenly considered fraudulent transaction as genuine transactions. TN is True Negative shows the number of fraudulent transactions correctly identified as frauds. Achieving the highest fraud catching and lowest false alarm rates are the most important tasks of the support vector machine models. The True Positive (TP) and False Positive (FP) rates are calculated using the following equations:

$$TP_{rate} = \frac{TP}{TP + FN} \tag{3.19}$$

$$[FP_{rate} = \frac{FP}{TN + FP} \tag{3.20}$$

TP_{rate} denotes the ratio of positive class that was correctly classified. FP_{rate} denotes the ratio of negative cases that were incorrectly classified as positive. Accuracy denotes the ratio of the total number of transactions that were correctly

classified. The accuracy of the classifier is calculated as shown below:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{3.21}$$

$$\text{Error Rate} = \frac{FP + FN}{TP + FP + FN + TN} \tag{3.22}$$

Chapter Four

Results & Discussion

4.1 Introduction

This chapter presents a detailed computational work on the data churned from running the study algorithms on the available database using the statistical methods discussed in the previous chapter. This chapter also explains the reasoning behind every method used and the outcome of the analysis.

4.1.1 Data Preparation

The data represents transactions of credit cardholders where there were 20 frauds out of 6000 credit card total transactions.

Table 4.1: Distribution of dataset

Type of transaction	Number of transactions	Percentage of transactions
Genuine	5980	99.67%
Fraudulent	20	0.33%
Total	6000	100.00%

From Table 4.1 the probability of the genuine transactions is 99.67% with the fraudulent transactions having a probability of 0.33%. The linear support vector machine algorithm produced a probability value of 0.17% in classifying the fraudulent transactions. However, the radial basis support vector machine algorithm produced a probability value of 0.49% in classifying the fraudulent transactions. On the other hand, the polynomial support vector machine algorithm produced a probability value of 0.47% in classifying the fraudulent transactions. From Table 4.2, the polynomial support vector machine algorithm has classified the fraudulent transactions better in terms of probabilities than the other support vector machine algorithms since its probability value of 0.47% is closer to the prior probability of 0.33% as shown in Table 4.1

Table 4.2: Distribution of probabilities of the SVM algorithms in percentages

Type of transaction	Linear kernel	Polynomial kernel	Radial Basis kernel
Non - Fraudulent	99.83	99.53	99.51
Fraudulent	0.17	0.47	0.49
Total	100	100	100

The classification accuracy of the testing data is a measure to calculate the ability of the classifier to detect and identify fraudulent transactions. The testing data used to assess and evaluate the efficiency of the proposed classifiers were taken only from credit card transactions from `http://www.kaggle.com`. For the purpose of statistical and machine learning classification tasks, a confusion matrix also known as an error matrix, is a specific table layout that allows the visualization of the performance of a supervised learning algorithm.

Table 4.3: Confusion Matrix of the Linear SVM

		Predicted	
		Positive	**Negative**
Actual	Positive	5999	8
	Negative	1	12

Table 4.3 shows the confusion matrix when the linear SVM algorithm is used for classification. In judging the performance of the classifiers obtained in the methods discussed, the following measures are calculated; that is Accuracy, Sensitivity, Specificity and Precision with the concept of True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). From Table 4.3, we deduce the following results: $TP = 5999$ which is the number of genuine transactions that have been classified as genuine. $TN = 12$ which is the number of fraudulent transactions that have been classified as frauds. $FP = 1$ which is the number of non - frauds that have been classified as frauds. $FN = 8$ which is the number of frauds that have been classified as non - frauds.

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{5999}{6007} = 99.87\%$$

This means that the proportion of genuine transactions that have been correctly classified using the linear support vector machine algorithm as genuine is 99.87%.

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{12}{1 + 12} = 93.31\%$$

This means the proportion of fraudulent transactions that have been correctly clas-

sified as fraudulent using the linear support vector machine algorithm is 93.31%.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} = \frac{5999 + 12}{5999 + 8 + 1 + 12} = 99.85\%$$

This means that the proportion of all the transactions that were correctly classified as genuine or fraudulent using the linear support vector machine algorithm is 99.85%.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{5999}{5995 + 1} = 99.98\%$$

Table 4.4: Confusion Matrix of the Polynomial SVM

		Predicted	
		Positive	**Negative**
Actual	Positive	5999	5
	Negative	1	15

Table 4.4 shows the confusion matrix when the polynomial SVM algorithm is used for classification. From Table 4.4, we calculate the performance measures as follow:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{5999}{5999 + 1} = 99.98\%$$

This means that the proportion of genuine transactions that have been correctly classified using the polynomial support vector machine algorithm as genuine is 99.98%.

$$\text{Specificity} = \frac{TP}{FP + TN} = \frac{15}{1 + 15} = 93.75\%$$

This means the proportion of fraudulent transactions that have been correctly classified as fraudulent using the polynomial support vector machine algorithm

is 93.75%.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{5999 + 15}{5999 + 15 + 1 + 5} = 99.90\%$$

This means that the proportion of all the transactions that were correctly classified as genuine or fraudulent using the polynomial support vector machine algorithm is 99.90%.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{5999}{5995 + 1} = 99.98\%$$

Table 4.5: Confusion Matrix of the RBF SVM

		Predicted	
		Positive	**Negative**
Actual	Positive	5995	8
	Negative	5	12

Table 4.5 shows the confusion matrix when the RBF SVM algorithm is used for classification. From Table 4.5, we calculate the performance measures as follow:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{5995}{5995 + 8} = 99.87\%$$

This means that the proportion of genuine transactions that have been correctly classified using the RBF support vector machine algorithm as genuine is 99.87%.

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{12}{12 + 5} = 70.59\%$$

This means the proportion of fraudulent transactions that have been correctly clas-

sified as fraudulent using the RBF support vector machine algorithm is 70.59%.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{5995 + 12}{5995 + 12 + 5 + 8} = 99.78\%$$

This means that the proportion of all the transactions that were correctly classified as genuine or fraudulent using the RBF support vector machine algorithm is 99.78%.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{5995}{5995 + 5} = 99.92\%$$

Table 4.6 contains the performance measures when the three classifiers (linear SVM, polynomial SVM, RBF SVM) are used for classification.

Table 4.6: Average performance analysis of SVM classifiers in percentages

Description	Accuracy	Sensitivity	Specificity	Precision
Linear	99.85	99.87	93.31	99.98
Polynomial	99.90	99.98	93.75	99.98
RBF	99.78	99.87	70.59	99.92

From the accuracy values shown in Table 4.6, it is clear that the support vector machine using the three kernel functions have all performed very well even though the support vector machine using the polynomial kernel function seems to have performed better than the other kernel functions. This conclusion is supported by the probability results discussed earlier.

Plate 4.1: Confusion matrices of the ANN algorithm

From Plate 4.1, we calculate the performance measures as follow:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{5979}{5979 + 0} = 100.00\%$$

This means that the proportion of genuine transactions that have been correctly classified using the ANN algorithm as genuine is 100.00%.

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{20}{20 + 21} = 48.78\%$$

This means the proportion of fraudulent transactions that have been correctly clas-

sified as fraudulent using the ANN algorithm is 48.78%.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{5979 + 20}{5979 + 20 + 21 + 0} = 99.65\%$$

This means that the proportion of all the transactions that were correctly classified as genuine or fraudulent using the ANN algorithm is 99.65%.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{5979}{5979 + 21} = 99.65\%$$

Table 4.7: Confusion Matrix of the k - NN algorithm

		Predicted	
		Positive	**Negative**
Actual	Positive	6000	8
	Negative	0	12

Table 4.7 shows the confusion matrix when the KNN algorithm is used for classification. From Table 4.7, we calculate the performance measures as follow:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{6000}{6000 + 8} = 99.87\%$$

This means that the proportion of genuine transactions that have been correctly classified using the KNN algorithm as genuine is $99.87.00\%$.

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{12}{0 + 12} = 100\%$$

This means the proportion of fraudulent transactions that have been correctly clas-

sified as fraudulent using the k - NN algorithm is 100.00%.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{6000 + 12}{6000 + 12 + +8} = 99.87\%$$

This means that the proportion of all the transactions that were correctly classified as genuine or fraudulent using the KNN algorithm is 99.87%.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{6000}{6000 + 0} = 100\%$$

In comparing the performance of different models, the F1 - score is used. Hence the greater the F1 - score the better the model.

$$\text{F1 - score} = \frac{2 \times sensitivity \times precision}{sensitivity + precision}$$

The F1 - score value for the polynomial SVM algorithm

$$= \frac{2 \times 99.98 \times 99.98}{99.98 + 99.98} = 99.96\%$$

The F1 - score value for the KNN algorithm

$$= \frac{2 \times 99.87 \times 100}{99.87 + 100} = 99.93\%$$

The F1 - score value for the ANN algorithm

$$= \frac{2 \times 100 \times 99.65}{100 + 99.65} = 99.81\%$$

Table 4.8 contains the F1 - scores of the three classifiers (polynomial SVM, KNN

and ANN) algorithms.

Table 4.8: Percentage F1 - Scores

Description	Polynomial SVM	KNN	ANN
F1 - Score	99.98	99.93	99.83

From Table 4.8, the polynomial support vector machine algorithm has the highest F1 - score of 99.96%. This therefore suggests that the polynomial support vector machine has performed better in classifying the data than any of the other algorithms used in the analysis.

Chapter Five

Summary, Conclusion & Recommendation

5.1 Summary

From the distribution of the data, actual probability of the fraudulent transactions is 0.33%. Using the SVM classifiers, The linear SVM algorithm classified the fraudulent transactions with probability of 0.17%. The accuracy score is 99.85%, the sensitivity score is 99.87%, the specificity score is 93.31% with a precision value of 99.98%. The polynomial SVM algorithm classified the fraudulent transactions with probability of 0.47%. The accuracy score is 99.90%, the sensitivity score is 99.98%, specificity score is 93.75%, precision score is 99.98% with an F1score of 99.98. The RBF SVM algorithm classified the fraudulent transactions with probability of 0.49%. The accuracy score is 99.78%, the sensitivity score is 99.87%, specificity score of 70.59% with a precision score of 99.92%. The ANN algorithm classified the fraudulent transactions with probability score of 0.30%. The accuracy score is 99.65%, sensitivity score of 100.00%, specificity score is 48.78%, precision score is 99.65% with an F1score of 99.83%. The KNN algorithm produced an accuracy score of 99.87%, sensitivity score of 99.87%,

specificity score of 100.00% with an F1score of 99.93%.

5.2 Conclusion

This study presents and demonstrates the advantages of support vector machine, artificial neural network and the k - nearest neighbor algorithms to the credit cards data for the purpose of reducing the bank's losses. Although all the three algorithms have performed very well in classifying the data into fraudulent and non - fraudulent transactions, the polynomial support vector machine algorithm classified the data with more accuracy than the other two algorithms signifying that it is the algorithm with the lowest error rate. However, the artificial neural network algorithm has classified the data very well into number of fraudulent transactions and non - fraudulent transactions thereby reducing drastically mis - classification and the generation of false alarms. In order to compare the performance of these algorithms, I calculated the True Positive, True Negative, False Positive and False Negative as quantitative measurements to evaluate and compare the performance of these algorithms. All the three algorithms have performed very well with respect to the data used in the study. With accuracy values of 99.90%, 99.65% and 99.87% respectively. With these high levels of accurate performances, it will go a long way to help the banks and other financial institutions to minimize their annual losses drastically that would have been incurred through credit card fraud activities if they incorporate these algorithms into their operational systems.

5.3 Recommendation

Although there are several credit card fraud detection techniques available today, none is able to detect all frauds completely when they are actually happening. They usually are able to detect them after the frauds have been committed. This happens because a minuscule number of transactions from the total transactions are actually fraudulent in nature. So we need a technology that can detect a fraudulent transaction when it is taking place so that it can be stopped then and there at a very minimum cost. The major task therefore is to build an accurate, precise and fast detecting systems for credit card fraud detection. The major draw back of all the techniques is that they are not guaranteed to give the same results in all environments. They give better results with a particular type of data and poor or unsatisfactory results with other types of data. A solution to this gaps will be by creating a hybrid of various techniques that are already used in credit card fraud detection to cancel out their individual limitations and then get enhanced performance. However, this requires official support from the credit card issuers.

Bibliography

Aha, D. W., Kibler, D., and Albert, M. K. (1991). Instance-based learning algorithms. *Machine learning*, 6(1):37–66.

Aleskerov, E., Freisleben, B., and Rao, B. (1997). Cardwatch: A neural network based database mining system for credit card fraud detection. In *Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr)*, pages 220–226. IEEE.

Axelsson, S. (2000). The base-rate fallacy and the difficulty of intrusion detection. *ACM Transactions on Information and System Security (TISSEC)*, 3(3):186–205.

Bolton, R. J. and Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical science*, 17(3):235–255.

Brabazon, A., Cahill, J., Keenan, P., and Walsh, D. (2010). Identifying online credit card fraud using artificial immune systems. In *IEEE Congress on Evolutionary Computation*, pages 1–7. IEEE.

Brause, R., Langsdorf, T., and Hepp, M. (1999). Neural data mining for credit

card fraud detection. In *Proceedings 11th International Conference on Tools with Artificial Intelligence*, pages 103–106. IEEE.

Chan, P. K., Fan, W., Prodromidis, A. L., and Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems and Their Applications*, 14(6):67–74.

Chaudhary, K., Yadav, J., and Mallick, B. (2012). A review of fraud detection techniques: Credit card. *International Journal of Computer Applications*, 45(1):39–44.

Chiu, C.-C. and Tsai, C.-Y. (2007). A k-anonymity clustering method for effective data privacy preservation. In *International Conference on Advanced Data Mining and Applications*, pages 89–99. Springer.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Delamaire, L., Abdou, H., and Pointon, J. (2009). Credit card fraud and detection techniques: a review. *Banks and Bank systems*, 4(2):57–68.

Elkan, C. (2001). The foundations of cost-sensitive learning. In *International joint conference on artificial intelligence*, volume 17, pages 973–978. Lawrence Erlbaum Associates Ltd.

Gadi, M. F. A., Wang, X., and Lago, A. P. d. (2008). Credit card fraud detection with artificial immune system. In *International conference on artificial immune systems*, pages 119–131. Springer.

Hussain, A. J., Knowles, A., Lisboa, P. J., and El-Deredy, W. (2008). Financial time series prediction using polynomial pipelined neural networks. *Expert Systems with Applications*, 35(3):1186–1199.

Lavrac, N., Kavsek, B., Flach, P., and Todorovski, L. (2004). Subgroup discovery with cn2-sd. *J. Mach. Learn. Res.*, 5(2):153–188.

Lee, W., Stolfo, S. J., Chan, P. K., Eskin, E., Fan, W., Miller, M., Hershkop, S., and Zhang, J. (2001). Real time data mining-based intrusion detection. In *Proceedings DARPA Information Survivability Conference and Exposition II. DISCEX'01*, volume 1, pages 89–100. IEEE.

Panigrahi, S., Kundu, A., Sural, S., and Majumdar, A. K. (2009). Credit card fraud detection: A fusion approach using dempster–shafer theory and bayesian learning. *Information Fusion*, 10(4):354–363.

Pimentel, M. A., Clifton, D. A., Clifton, L., and Tarassenko, L. (2014). A review of novelty detection. *Signal processing*, 99:215–249.

Srivastava, A., Kundu, A., Sural, S., and Majumdar, A. (2008). Credit card fraud detection using hidden markov model. *IEEE Transactions on dependable and secure computing*, 5(1):37–48.

Sun, L. M., Chiu, H.-W., Chuang, C. Y., and Liu, L. (2011). A prediction model based on an artificial intelligence system for moderate to severe obstructive sleep apnea. *Sleep and Breathing*, 15(3):317–323.

Vatsa, M., Singh, R., and Noore, A. (2005). Improving biometric recognition

accuracy and robustness using dwt and svm watermarking. *IEICE Electronics Express*, 2(12):362–367.

Wheeler, R. and Aitken, S. (2000). Multiple algorithms for fraud detection. In *Applications and Innovations in Intelligent Systems VII*, pages 219–231. Springer.

Yu, W.-F. and Wang, N. (2009). Research on credit card fraud detection model based on distance sum. In *2009 International Joint Conference on Artificial Intelligence*, pages 353–356. IEEE.

Zhang, Y., You, F., and Liu, H. (2009). Behavior-based credit card fraud detecting model. In *2009 Fifth International Joint conference on INC, IMS and IDC*, pages 855–858. IEEE.