T.C.

ALTINBAS UNIVERSITY

Graduate School of Science and Engineering

Information Technology

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING METHODOLOGY

HAMZAH ALI SHUKAUR ALMARSOOMI

Master Thesis

Supervisor

Asst.Prof. Dr. Sefer Kurnaz

Istanbul,2019

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING METHODOLOGY

by

Hamzah Ali Shukur Almarsoomi

Information Technologies

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science

ALTINBAŞ UNIVERSITY
2019

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of .....

Academic Title Name SURNAME Co-Supervisor	Academic Title Name SURNAME Supervisor

Examining Committee Members (first name belongs to the chairperson of the jury and the second name belongs to supervisor)

Academic Title Name SURNAME	Faculty, University	_____
Academic Title Name SURNAME	Faculty, University	_____
Academic Title Name SURNAME	Faculty, University	_____
Academic Title Name SURNAME	Faculty, University	_____
Assoc. Prof. Name SURNAME	Faculty, University	_____

I certify that this thesis satisfies all the requirements as a thesis for the degree of .....

Approval Date of Graduate School of
Science and Engineering: ____/____/____

_____
Academic Title Name SURNAME
Head of Department

_____
Academic Title Name SURNAME
Director

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Hamzah Ali Almarsoomi

DEDICATION

I would like to dedicate this work to my first teacher, my mother, my first supporter and role model, my father and my companion throughout the journey. Without you, this dream would never come true and to my brother and my sister who stood with me in order to achieve my dream.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all the instructors that have taught me more than just science, especially my supervisor Asst.Prof.Dr.Sefer Kurnaz, for all the time, support and guidance provided to me along the journey to accomplish this work. Thank you all for all the knowledge and advice that made me overcome all the difficulties that I have faces.

ABSTRACT

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING METHODOLOGY

Almarsoomi, Hamzah Ali Shukur

M.Sc. Information Technologies, Altınbaş University,

Supervisor: Asst. Prof. Dr. Sefer KURNAZ

Date: Mar/2019

Pages: 65

In USA,credit card fraud occurrences rose sharply in 1998 causing SI47 million in losses. To address this problem,financial institutions (FIs) are employing preventive measures and fraud.detection systems one of which is called DS. Although FDS has shown good results in reducing fraud, the majority of cases being flagged by this system are False Positives resulting in sub-stantial investigation costs and cardholder inconvenience.The possibilities of enhancing the current operation by introducing a post processing system constitute the objective of this research.The data used for the analysis was provided by one of the major Canadian banks Based on variations and combinations of features and training class distributions,different sets of experiments were performed to explore the influence of these parameters on the performance of The prototype developed. The results indicate that the employed approach has a

very good potential to improve on the existing system.However further research is required including the development of prototype systems which should be enhanced by more extensive and informative data.

Keywords: Financial institutions , Fraud Detection Systems , Prototype Systems.

TABLE OF CONTENTS

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABBREVIATIONS

ABM	:	Automated Banking Machine
CID	:	Card Identification Device
CI	:	Card Issuers
CVV	:	Card Verification Value
CVC	:	Card Verification Code
CBA	:	Canadian Bankers Association

1. INTRODUCTION

The main of credit shipset was started at the end of the first World War in the United States (U.S.). In 1956, Bank of America which is named toduy (now VISA) opened the business followed by MasterCard. In 1968, 4 Canadian banks launched VISA credit card to meet the business demand.

1.1 PROBLEM DEFINITIONS

In the period of credit cards, fighting fraud was rather straightforward. Each week a bulletin known as a 'hot list', was passed resolute the merchants. This bulletin noncommissioned the varieties of lost or purloined cards in order that the merchants were ready to check the client MasterCard number against these numbers. With the blowup of the MasterCard business, criminals have devised varied ways that to urge about ahead security devices, like magnetic stripes and holograms. Cardholders are the prominent victims of fraud occurrences, as they are going to buy the value of fraud injuries sustained by either card issuers or merchants. To atone for these losses card issuers, raise the interest rates or seasonal fees and merchants raise the worth of their wares. Fraud investigation is a difficult task and FIs are reluctant to block an account without making sure that the transaction is indeed fraudulent. Very often an unusual transaction is legitimate and issuers are anxious not to inadvertently off end a cardholder by acting too hastily and blocking her/his account especially in cases where the fraud officer is unable to find the cardholders and verify the transactions with them.FDS has shown good results in detecting fraudulent transactions, however, the majority of transactions (approximately 90%) being flagged by this system as potentially fraudulent are in fact legitimate. It should be noted that although fraud analysts based on their experience and evaluation of the customer's history might come to the conclusion that the activity of the flagged account is legitimate bank policy requires them to call every individual cardholder for the verification of transactions [12]. The process of calling cardholders results in three major problems:

1. Not all the suspicious transactions are necessarily fraudulent. This type of error is referred as false positive (FP) which means that the case was not fraud although it was flagged as being potentially fraudulent. The process of confirming every transaction that deviated from the cardholder's usual behavior results in potential.
2. The costs associated with investigating a large number of false positives are very high.

3. Currently, a substantial amount of time is being spent on investigating a large number of legitimate cases (FPs). If the number of investigations on FPs could be lowered down, fraud analysts can spend more time on real fraud cases, preventing more losses to the industry.

1.2 CONTRIBUTION OF THESIS

As pointed out earlier, the verification of suspicious transactions with the cardholder is a major part of fraud investigation and cannot be eliminated. Therefore, any solution that refines the investigation selection process by reducing the number of unnecessary calls is welcomed by the FIs. Collaboration with one of the major INTERNATIONAL banks was established to examine the potential ways of enhancing the current system. Based on information obtained from this bank, for the current threshold, FDS flags close to 50,000 accounts per month all across Canada. The main objective of this research is to improve the process of personal follow up on a large number of suspicious transactions. Moreover, the current FDS threshold can also be lowered and a number of fraudulent cases, being missed under this level, can be detected. As a result, the fraud is discovered earlier, and the overall losses may be reduced.

1.3 THESIS ORGANIZATION

The thesis organization fall into seven chapters, in Chapter 2 gives an overview of the history of credit cards, the transaction and authorization processing operation of the FIs and the ways that this convenient method of payment has been endangered by criminals. It proceeds to explore the types of fraud and concludes with statistics and some facts regarding credit card fraud occurrences in Canada. In Chapter 3 introduces the existing fraud solution approaches and gives a brief introduction to the existing Fraud Detection Systems (FDS). It touches on neural network technology, its advantages and disadvantages and briefly describes its application to credit card fraud detection. It further elaborates on fraud investigation processes and the associated issues. In Chapter 4 display the notion of classification and the principal strands of research in this area. It gives an overview of learning systems, their requirements, and describes learning decision trees and their applications. In Chapter 5 provides a detailed explanation on how data was acquired, and the steps required for processing the data to make it ready for the analysis. It introduces the learning software and touches on its capabilities. It further elaborates on the data set and class distribution designs for training and testing sets, and the variations of experiments performed. In Chapter 6 shows

the outcomes of the analyses an the effects of the training class distributions on the results. It proceeds to compare the performance of different classifiers with each other, and presents the efficient classifier based on the criteria defined. It also examines the evaluation prediction of two classifiers on the prediction of new cases. In Chapter 7 presents the conclusions of this research and offers suggestions for further study.

2. LITERATURE REVIEW

2.1 CONVENIENT METHOD OF PAYMENT

A credit card is a special product with the following characteristics [21|: It provides millions of people around the world with the opportunity to purchase goods and services with access to credit for up to 51 days, depending on the posted date of the purchase, at no cost provided that the amount owing is paid back by the statement due date. Cardholders do not have to put up collateral against the amount they spend, therefore, it is unsecured. In North America, credit cards are widely used in purchasing goods and services. The main reasons for this popularity are:

1. The existence of a widespread point of sale (POS) network.
2. Reducing the risk of carrying cash and the advantage of several weeks of free credit plus optional services and benefits such as Air Miles, free insurance plans, and a number of other rewards.
3. Security of funds, that is, in case of card loss or theft, the cardholder's liability at the most is $50 provided that the cardholder reports a lost or stolen card in a timely manner.

The credit card system facilitates commercial transactions and provides profits for the Participating parties. The source of income for card issuers (CIs) may come from: (1) Merchant user fees, (2) cardholder user fees, and (3) interest charged on unpaid balances. In purchasing goods and services the buyer pays for a purchase by using a line of credit from the credit card issuer (Cl). The Cl pays the seller for the purchase, and the buyer then pays the balance on the credit card back to the Cl. Since the claim presented in payment is considered a liability of the credit card.issuer, this type of transaction transfers much of the Risk of insufficient funds in the transaction from the seller to the credit card issuer. order to make up for these losses, CIs determine annual fees and interest rates based on the Unrecoverable amount of money incurred by these losses. It is worthwhile to point out that Most of the CIs are FIs even though there are many which are not. In this work, we use FI Owned credit card operation as our "laboratory".

2.2 CREDIT CARDS TRANSACTION PROCESS

The following information on the credit card transaction processes was collected through personal meetings with the staff of the sponsoring financial institution and review of a Master thesis on this subject.

2.2.1 Parties Involved in a Transaction

Four parties are involved in processing a credit card transaction: (I) the cardholder, (2) the merchant, (3) the Financial Institution (FI), and (4) the VISA center.

1. The card holder uses the card for a purchase and provided that the statement amount is paid back by the due date, interest charges will not occur.
2. The merchant by accepting the card for payment, has the advantage of security of payment by the FIs.
3. The FIs issue the cards, settle other FI's cardholder and merchant transactions with VISA, process the incoming transactions and provide the cardholders with monthly statements.

Normally for every transaction, one or two FIs are involved: the cardholder's and the merchants. The cardholders of one FI might go to the merchants of the same FI or another FI. Therefore, depending on the situation, the FI could assume two roles, being an agent for both the cardholder and the merchant or being an agent for either one of them. Hence, the transaction processing system must be able to separate the incoming transactions of a particular FI from the other FIs and route each transaction to the appropriate place for authorization and record keeping.

2.2.2 Overview of Transaction Processing Flow

In purchasing goods or services through credit cards, in on-line processing systems, the authorization is the essential element of the transaction processing system. The authorization process is the first level of protection against fraudulent activities and it also maintains control over the cardholder's credit limit. It should be noted that the authorization is kept as a temporary file for up to five days and when the transaction is recorded in the cardholder file, the cardholder account balance is updated. In addition to online authorizations, there are merchants who have floor limits. If the amount of the transaction is below that limit, the authorization does not need to go through the FI's system, and the merchant has the right to authorize the transaction locally. In fact, due to a widespread POS network in North America

many merchants have ’zero’ floor limit and almost every transaction has to be authorized on-line by the related FI. The authorization process begins when a cardholder uses his/her card for a transaction. The POS machine reads the magnetic stripe embossed on the back of the card which encodes the card holder’s name, account number, credit limit, and the expiry date. The authorization is completed when the transaction is approved, and the cardholder signs the transaction slip. The next step is the submission of transaction slips to the FI which either is done electronically or manually. In electronic transfer, the POS machine keeps track of all the authorizations and sends them electronically to the FI. In this case the merchants do not need to submit the transaction slips. The manual option is when the merchant sends the actual slips to the FI and the FI’s operators will enter the records manually into the system. In both cases after the submission of transactions, the FI credits the merchant’s account by that amount. To handle a large number of cardholders’ monthly statements, FIs have set up several billing cycles during the month. A certain number of cardholders are associated with each of these cycles and the date for each billing cycle is different from the other cycles. At the end of each cycle a new statement is processed and mailed to the cardholder of that cycle along with a due date for payment. The statement contains information on the transactions such as the date, the reference number, the description, and the amount. Other information such as the previous and the new balance, the minimum payment, and the available credit is also included. The cardholder is required to pay the total or part of the balance. If the balance is paid back in full there are no interest charges. If the cardholder’s payment is less than the minimum amount, the credit rating of the cardholder could be affected and the cardholder may be considered delinquent. Another type of transaction possible by credit cards, is obtaining cash advances. In this type of transaction, the interest is charged from the day when the money is withdrawn even though the balance is paid back in full on the statement due date. For handling a huge number of daily transactions, FIs and VISA have implemented a real time, non-stop system of computer hardware and software. This system includes the communications network among the FIs and the VISA network as well as handles the data processing and the record keeping tasks. Figure 2-1 depicts an overview of the VISA transaction processing system. The main components of the system are described below.

2.2.2.1 The card

A credit card is a standard plastic card with a magnetic stripe on its back which is read by a POS machine at the point of purchase. The front side of the card has the cardholder’s name, account number, the expiry date and a hologram. Very often the credit limit on gold

and platinum cards is much higher than the classic types but these cards have annual fees as well. Services and rewards, such as insurance coverage for car rental, are mostly associated with these types of cards.

2.2.2.2 The swipe machine

POS terminals or swipe machines are very common in North America and are used for online authorization. After swiping the card through the machine and entering the amount of purchase on the keypad, the POS terminal reads the card's magnetic stripe information and places a call to the merchant FI's computer. This information, along with the merchant number, is transmitted via a modem, to the on-line authorization system. This service is typically processed by non-stop Tandem computers.

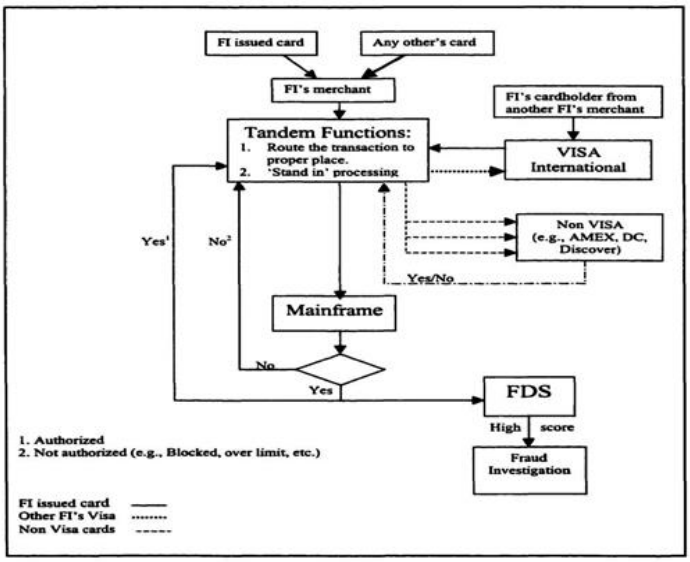

Figure 2.1: Overview of VISA Processing Transaction System.

2.2.2.3 The tandem

The Tandem is a non-stop computer used to process all the incoming electronic transactions regardless of the merchant's institution and country. Every FI has its own Tandem which is connected to the VISA network. The functionality of the Tandem is summarized below:

1. Keeping a record of all incoming transactions for further referral in case of any system malfunction. The incoming transactions are categorized as follows:

A. transaction by the merchant and the cardholder from the same FI.

B. transaction by the FI's merchant with a cardholder from another FIs. These transactions will be routed to VISA network and from there they will be sent to the cardholder FI's Tandem.

C. transaction by the FI's cardholder with another FI merchant. These transactions are sent to the VISA network and from there they are routed to the cardholder's FI Tandem for authorization.

2. There are occasions when the FI's mainframe is not able to do the authorization Processing due to: (I) a system breakdown, (2) when the FI's computer system is down for different reasons (e.g., maintenance). In these circumstances, the Tandem does 'stand-in' authorization processing, that is, it authorizes a transaction on behalf of the Mainframe. This process is described below:

The Tandem authorizes the transactions based on a 'negative file' and an assigned floor limit. Negative file includes all the card numbers that have been considered fraudulent internationally. This list is provided by Visa international and is updated quite frequently with the occurrence of new fraud cases. Before any authorization, the Tandem checks that the card number is not on that list. The Tandem does not have cardholder's account information and, therefore, it cannot do any credit limit checking for the cardholder's account but there is a set credit limit for the incoming transactions that the Tandem will check and will not exceed. When the cardholder FI's mainframe becomes available again, the Tandem will send the approved authorized transactions to the mainframe either in real time or as a batch file, depending on the circumstances.

2.2.2.4 The mainframe

The Tandem sorts the incoming transactions and transmits only those transactions which are from the FI's cardholders or merchants to the FI's mainframe computer for authorization. The mainframe, as the main component of the system, processes all the incoming authorization requests. For authorization the mainframe performs a series of checks to ensure that the customer is eligible for making purchases. Some of these checks are listed here [8]:

A. Card Expiry Date: If the card is expired the authorization is not allowed and the transaction is declined.

B. Excessive Authorizations: Under the normal situations a client will not exceed a certain number of will either be declined (D) or referred (R) (i.e., referring the case to the FI staff) transactions in a 24-hour period. This check limits the number of authorizations that a customer can do during that period. If an account goes over the allowed number for the day, the authorization.

C. Blocked: Block codes are used to put conditions on accounts. An account is checked for being fraudulent, delinquent or blocked. If any of these checks are positive, the authorization will

D. Credit Limit Check: This check verifies that the cardholder has not exceeded his/her credit limit. If the sum of the current transaction amount and the current balance is under the credit limit, the authorization will be approved, otherwise it will be declined.

In the checking process, if one of the required checks for the transaction fails, the authorization is declined, and this refusal will be sent to the Tandem and from there it will be sent back to the merchant. When the transaction passes all the required checks, the approved authorization goes back to the Tandem and from there, is sent back to the merchant. To make all these activities happen, FIs have implemented several sophisticated software packages. The databases required to track the aforementioned activities are as follows:

1. Merchant File: Information on the FI's merchant are included in this database.
2. Cardholder File: Information on the FI's cardholder account such as name, address, account number, current balance, credit limit, expiry date, and so on are contained in this file.
3. Authorization Log: All the authorized transactions done by the FI for its own cardholders are included in this file.
4. Posted TX file: This file keeps a record of all the transactions that have been received from the Tandem but have not yet been posted to the cardholders' monthly statement.
5. Statement File: At the end of the cardholder's cycle, the accumulated transactions in the posted TX file will be sent to this file and the monthly statement for the cardholder is printed out of this file.

When an authorization is approved the account's available credit and amount/number of authorizations are updated and the mainframe sends this information to the authorization log database and the posted TX file. At the end of each cycle date, all the posted transactions of each account from the posted TX file will be sent to the statement file processor. This is used in printing out the monthly statements of the cardholders. When the cardholder pays the total or the minimum due amount, the cardholder's file is updated by this payment and the current balance is adjusted. To save computer disk space, the statement file keeps a record of the last three statements and by the production of a new statement the oldest statement is archived.2.3 Credit Card Fraud

Plastic card-based payment systems are booming and being used more extensively by organizations and individuals. Obviously, industries with this pace of growth are vulnerable to attacks by fraudsters. In one survey [26] conducted in the United State (U.S.) in 1993, a group of 14 credit card fraudsters admitted to employing over 100 different ways of using credit cards to obtain funds illegally.

2.3.1 Fraud Schemes

Unauthorized use of credit cards for acquiring goods or services is fraud. Visa and MasterCard constitute about 65 percent of all outstanding revolving credit worldwide and the substantial number of fraud occurrences is centered on one or both of these cards [25]. Most credit card fraud schemes fall into the following categories [CBA99b]:

1. Lost /Stolen
2. Never Received Issued (NRI) (Mail theft)
3. Counterfeit
4. Telemarketing and mail-order
5. Fraudulent applications

2.3.1.1 Lost and stolen

Lost and stolen cards account for the majority of fraud cases. Fifty five percent of Visa and forty nine percent of MasterCard losses are based on lost/stolen cards.The average loss.incurred.by.this.kind.of.fraud.is.$700.[2].When.acard.is.lost.or.stolen.the.opportunity.for. fraud starts. Workplaces glove compartments of cars and sporting facilities are the main sources of stolen cards. Very often these losses are caused by a relative or friend's unauthorized use of the card without the cardholder's knowledge [16]. Sometimes cardholders might sell their card to criminals, then report the card as lost or stolen or they might do shopping and then repudiate the event and report the card as lost or stolen.

2.3.1.2 Never received issued (NRI)

An average of 439,000 new, renewal and replacement cards are mailed every day. Never-received cards are cards being stolen from the mail, either internally or externally, while in transit from the card issuer to the legitimate customer. The card may be used and then be sold on the black market [ANON98a|. The average losses for this type of fraud are significant because the cardholder is not aware of the theft and by the time the fraud is detected, a

substantial amount of purchases has been made. Very often, only when the cardholder receives her/his monthly statement, does s/he realize that the card has been stolen in the mail. Visa's never-received card losses leapt 68 percent in 1997. The average losses from a never-received cards are about $1,500, double that of a lost or stolen card [ANON98a|. One of the new ways to prevent this type of fraud is to send the card to the cardholder as a worthless piece of plastic (electronically blocked). On the receipt of the card, the customers have to call the bank to activate their card.

2.3.1.3 Counterfeit

The fastest growing type of bank card fraud is the illegal counterfeiting of credit cards mainly Visa and MasterCard. By employing new technologies criminals are able to produce exact replicas of existing cards. The average reported losses due to this type of fraud are higher than any other fraud category estimated.at.about.$4,500.[2].

2.3.1.4 Telemarketing and mail-order fraud

There are occasions when a fraudulent merchant or telemarketer calls to sell a non-existent product over the phone and by acquiring the cardholder's card number processes a fraudulent charge against the account. It should be noted that this type of fraud, due to customer awareness is on the decline.

2.3.1.5 Fraudulent applications

In this kind of fraud, fraudsters provide FIs with false information and identities to acquire a credit card illegally. Unlike stolen cards these cards are not signed and it takes longer time before the fraud is detected. This kind of fraud, due to the awareness of the FIs, is on the decline

2.3.2 New Technologies and Card Counterfeiting

Card counterfeiting, in terms of frequency and severity, is on the rise. The basic principle underlying this kind of fraud, is an account number which could be obtained from different sources such as legitimate records made in hotels, restaurants, retailers, discarded drafts or computer software. In order to issue credit cards, financial institutions generate a series of numbers. From these numbers, a certain number (e.g., 500 of them) may be selected by a process known as skipping and being used for issuing credit cards. To defraud the FIs,

fraudsters may use an account generating software such as Creditmaster and Credit Wizard to determine the skipping code and reveal the valid credit card account numbers. Fraudsters may also use another type of software called Sniffers to find credit card numbers that individuals are sending online. This software searches the networks for 16 digit numbers, records them and sends them to the fraudster [2].

One of the latest methods of counterfeiting credit cards is 'skimming' or 'bit copying'. This is a process by which the magnetic stripe encoding from one card is copied to the stripe of another card. This method is one of the common methods of counterfeiting credit cards and is drastically on the rise in Canada. Public places such as particular restaurants and gas stations are major sources of these fraudulent activities [12]. The acquired number will then be embossed or encoded on a piece of plastic designed for this purpose. Whenever the number is embossed on an ordinary (blank) plastic card, it is called a white plastic fraud. When the number is embossed and/or encoded on an expired or stolen credit card (from which the original data have been removed) the result is an 'altered' or 'falsified' credit card. In the case of card alteration magnetic stripes are altered or manufactured using equipment that can be purchased at electronic stores. When the number is affixed onto a totally counterfeit credit card it is called a 'pure counterfeit' card [18]. Figure 2.2 depicts the card counterfeiting process schematically.

2.3.3 The Counterfeiting Process

To understand the complexity and the nature of card counterfeiting, it is important to introduce the methodology used by counterfeiters in their operations. With improvements in technology, Counterfeiting a credit card is often done by using desktop computer systems.

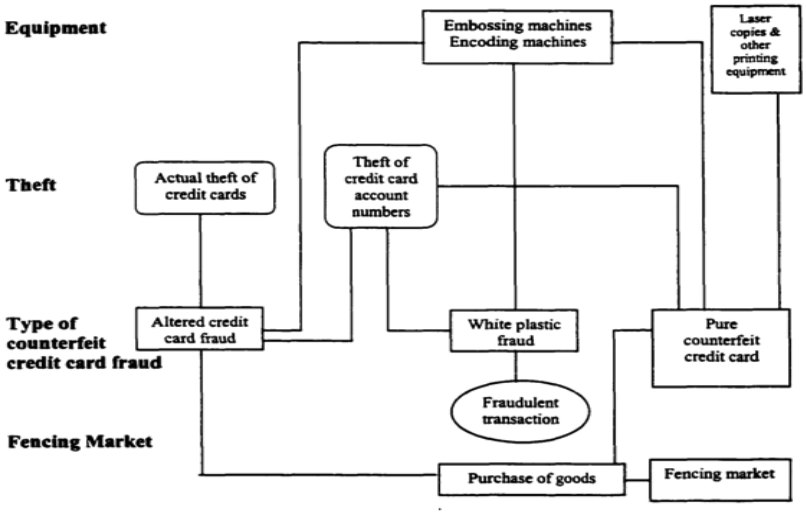

Figure 2.2: Fraudulent Transactions Using Counterfeit Cards [MAT197].

With peripherals including embossers, laminators, and tipping foil in order to produce a more realistic looking card complete with a hologram and fully encoded magnetic strip. Often the examination of the hologram is the key to the identification of a counterfeit card. On the legitimate cards, the hologram is embedded in the plastic at the time of manufacturing whereas counterfeit credit cards commonly contain a hologram affixed to the top of the card rather than embedded in the card. Thus, it can be seen or felt to rise slightly above the card face upon close examination [25]. The magnetic stripes and holograms used to counterfeit bank cards have a distinct sub- market within the criminal communities. Smugglers bring holograms into the U.S. and Canada regularly. During April 1994, the Canadian Combined Forces Special Enforcement Unit arrested members of a group that produced approximately 300,000 counterfeit holograms of which 250,000 had already been distributed. Based on the reported figures and an estimated loss of $3,000 per card, Visa and MasterCard anticipated losses of $750 million incurred by this organized activity [25]. The card counterfeiting in Canada is mainly an organized crime activity. This criminal activity started in Vancouver where fraudsters imported the technology of pure counterfeit credit cards from abroad and then it spread to East frontiers, mainly Toronto [MATI971. In mid-December 1998, police discovered a factory in the Toronto area that could produce cards from any financial institution including foreign ones. Police arrested a group of criminals who were charged with the production of counterfeit credit cards and Canadian cash. The associated charges for this criminal activity was so high (307 credit card related charges) that police announced this operation as the largest one, ever happened in Canada. One of the major concerns is that this information can also be sold to overseas groups who then can produce more counterfeit cards, hi addition to the losses imposed on the industry, the money obtained can also be used to buy more sophisticated equipment in order to produce more counterfeit cards and to expand. criminal activities worldwide. [LEMA98].

2.4 CREDIT CARDS IN CANADA

There are over 600 institutions in Canada that issue VISA or MasterCard. Among these CIs, the number of major institutions that issue VISA or MasterCard are 18; ten banks, one trust company, three credit unions, and four other financial institutions. The other CIs are affiliated issuers, such as the Bay, General Motors, University of Toronto, Petro-Canada, Eaton's, Canadian Tire, and so on. The number of credit cards issued by Financial institutions (FIs), all across the country, is approximately 35.3 million [CBA99a|. Figure 2.3 and Table 2.1

illustrate general statistics on Visa & MasterCard cards in Canada, respectively. Figure 2.4 was plotted based on statistics obtained from Canadian Bankers Association (CBA) web site.

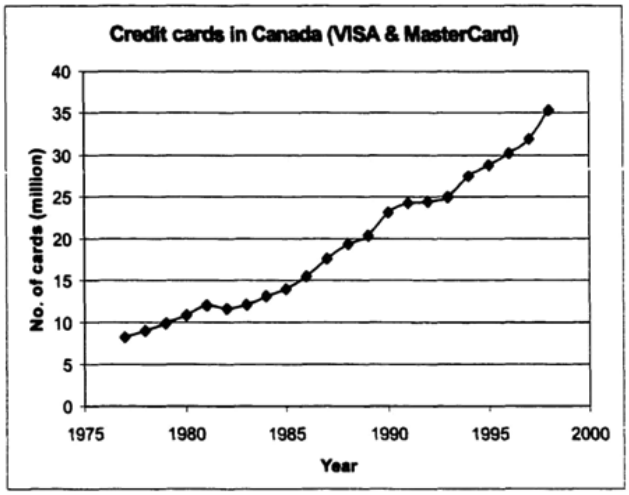

Figure 2.3: Credit Cards in Circulation in Canada [CBA99c].

Table 2.1: General Statistics on Credit Cards in Canada [CBA99a].

VISA & MASTERCARD	October 31,1998 (Fiscal year end)	October 31, 1997 (Fiscal year end)
Number of cards in circulation (million)	35.3	31.9
Outstanding balances ($ billion)	$23.9	$20.5
Retail sales volume ($ billion)	$84.1	$76.0
Delinquency ratios (90 days and over)	0.9%	0.9%
Sales slip processed (million)	1001.1	949.5
Average sales ($)	$90	$82.5
Sales and cash advance volume ($ billion)	$93.9	$84.3
Fraud losses ($ million)	$104.8	$88.0

2.4.1 Interest Rate Base

The number of outlets that accept VISA and/or MasterCard in Canada is approximately 620,000. Based on information obtained from CBA web site, average credit card interest rates

for standard cards, issued by Canada's six largest banks, have dropped by 3.4 percent since their peak in October 1990. Many banks now issue special low rate cards designed to benefit cardholders who usually do not pay off their balances every month. Compared to other credit cards types, low rate cards have significantly lower interest rates but slightly higher annual fees. On average, the interest on low rate card is more than six percent lower than the standard card rates. A number of factors such as cost of funds, losses due to fraud, level of fees and the volume of outstanding balances determine the base for interest rates. These factors are pointed out below [CBA99a]:

- Total losses due to credit card fraud was estimated at $147 million in 1998.
- As bankruptcies have become more acceptable, it has become more frequent, leading to increased losses in the credit card area.
- As the result of market pressures, the annual fees on standard credit cards have been eliminated.
- A higher percentage of Canadian cardholders pay off their balances in full or they have been carrying lower balances resulting in less interest income for the CIs from these sources.

Table 2.2: provides information on the interest rates in Canada.

Issuers/Cards	Interest Rate	Annual Fee
Banks (e.g., VISA, MasterCard)	8.99% - 18.9%	$0 - $39
Retailers, (e.g., Sears, Eaton's)	24% - 28.8%	$0

2.4.2 Statistics on Credit Card Fraud

Based on statistics reported by the CBA, credit card fraud occurrences rose sharply in fiscal 1998 (141,274) compared to 1997 (113,264). Based on the information obtained from this report, 34 percent of all credit card fraud occurrences and SO percent of the SI47 million written off in 1998 was due to counterfeit card fraud. This report also indicates that approximately SO percent of Canadian credit cards which were used fraudulently, were used outside of Canada. Figures 2.3 and 2.4 are plotted based on the statistics obtained from the CBA web site [CBA99b|. Figure 2.5 shows the number of cards used fraudulently in Canada. Figure 2.6 illustrates the statistics on different types of fraud in Canada.

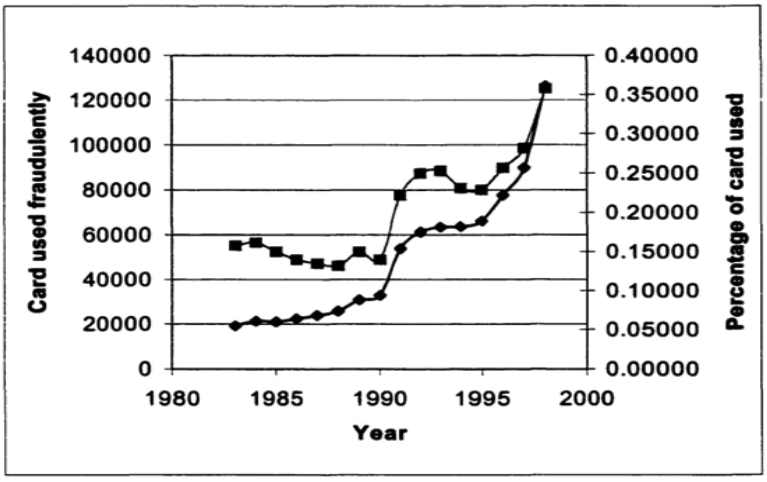

Figure 2.4: Cards reported fraudulently used in Canada [CBA99b].

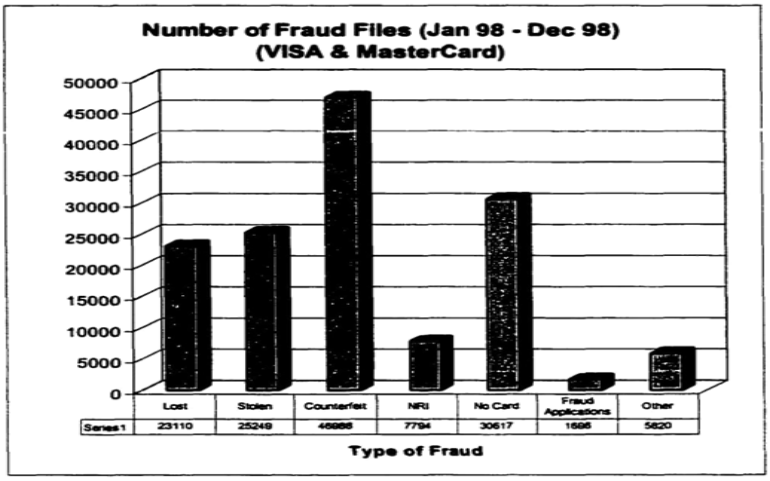

Figure 2.5: Statistics on different types of fraud in Canada [CBA99b].

2.5 SUMMARY

This Chapter presents an overview of the history of credit cards, the transaction and authorization processing operation of the FIs and the ways this convenient method of payment has been endangered by criminals. It proceeds to explore the types of fraud and concludes with statistics and some facts regarding credit card fraud occurrences in Canada.

3. FRAUD SOLUTION APPROACHES

3.1 MOTIVATION

FIs employ various technologies to detect and prevent credit card fraud. One of these, is special security numbers embedded in the magnetic stripe. The Card Verification Value (CW), Card Verification Code (CVC), and Card Identification (CID) are the security numbers being used by VISA, MasterCard, and American Express, respectively. This number, along with the account number and expiration date, forms an algorithm during the authorization process. If any part of this algorithm is missing or incorrect, the authorization at the point of sale (POS) will be declined [CBA98b]. For this reason, fraudsters not only need to have a valid account number but also need to know the mathematical formula used to create the code and the method of its encryption to be able to produce a counterfeit card. However, there are many situations where preventive techniques (e.g., holograms, validation codes such as CW) are not effective. For instance, in placing a telephone-order transaction or using the card over the Internet, these security features cannot be checked.

3.2 SMART CARDS

To address the problem, credit card manufacturers plan to employ a series of security features, most of which are designed to enhance customer identification and authorization requirements. Due to the shortcomings of holograms as a fraud preventive, the next generation of credit cards, called smart cards, has computer chips instead of holograms. Each card contains a microprocessor memory chip as well as data encoded on the magnetic stripe. For an authorization the cardholder is required to enter the personal identification number (PIN) encoded on the microchip. The industry foresees a time when bank customers will be able to use a single card to administer all their financial needs [ANON94b). Since the late 1980s, French banks with about 25 million smart cards in circulation, about half of the world's total of smart cards, have already made use of this technology and based on the reports their fraud volume has been cut drastically [13].

3.2.1 Implementation Issues in North America

Although the idea of smart cards seems very appealing, getting from the idea to practice in the North American market is another matter. Smart cards have become common in France and some other European and Asian markets due to the lack of a widespread communication

networks and the relatively costly telephone lines. Although by shifting to smart cards, the card issuing institutions could save more than a billion dollars per year, however, this conversion would be very costly. The main reasons are [13]:

1. Thanks to the fairly cheap telecommunication systems in North America, more than 90% of card transactions are authorized on-line. In the case of this implementation, POS terminals would need to be replaced or retrofitted for the current card use.
2. New cards have to be manufactured and distributed. The cost associated with issuing a smart card is up to six times higher than magnetic stripe cards. The cost is determined by the number of chips being mounted on the card.
3. An agreement, on a new system of fees, has to be established between the card issuers and their card organizations.

In the long term, however, smart cards will lead to significant cost savings. Although advancements in security technology are encouraging, smart cards are unlikely to become widespread until after the year 2000. In 1994, the cost of the infrastructure required to issue smart card worldwide was estimated at 7.4 billion dollars. Neither Visa nor MasterCard have yet been able.to justify these costs [1].

3.3 FRAUD DETECTION SYSTEMS

Along with the rise of credit card fraud, FIs are employing various methodologies and strategies to detect and prevent fraud. The main technologies used are pointed out in the following Sections.

3.3.1 Rule-Based Systems

Rule-based systems are computer programs, in the category of expert systems, consisting of a set of “If A then B” rules (where A is an assertion and B can be either an action or another assertion) designed to monitor transactions and flag unusual behavior such as high valued purchases or rapidly reaching the cardholder’s credit limit. The result is a list of suspicious transactions which will be handed in to fraud analysts for investigation.

3.3.2 In House Detection Software

There are occasions where the FIs devise their own systems based on their account histories and typical transactions. An example is the system used by American Express. It should be

noted that due to proprietary issues, there is not much information available on in house detection system; otherwise, it would be interesting to compare these systems to Visa system.

3.3.3 Neural Networks

Neural Networks (NNs) are a subdivision of Artificial Intelligence (AI) designed to address classification and pattern recognition problems. The term 'neural' is somewhat misleading. Also, the technology was inspired by the way neurons in the brain interact with each other, in reality there is no thinking in a neural network. Klimasauskas, Director of Financial Services at Neural Ware, a Pittsburgh based neural network vendor, has commented on this fact: "The important thing to realize is that neural networks, as a technology, have nothing to do with the brain. It is called neural because many of the techniques were first introduced by people who were studying the human brain but it is really a set of mathematical techniques for clustering information and finding curves for the data." [21].

3.3.3.1 The advantages of neural networks

Neural networks are able to capture associations or discover regularities within a set of variables. The application domain of NNs very much depends on the nature of the problem being modeled, but these systems are specifically suitable for domains where the relationships are dynamic and non-linear. In general, NNs are designed to address the following situations [21]:

- The number of variables or the volume of data is very diverse.
- The relationship among variables is inherently complex and cannot easily be identified.
- There is a need for modeling diverse behavior by finding patterns among cases.

3.3.3.2 The disadvantages of neural networks

While NNs have been used successfully for classification, they do suffer from the fact that the network is viewed as a black box and there is no explanation of the result. Due to the fact that the result is a completed network with layers and nodes linked together with non-linear functions whose relationship cannot easily be described, neural networks are generally difficult to understand. Moreover, they suffer from long learning times which become worse as the volume of data grows. Another major weakness of NNs is the lack of diagnostic help.

If something goes wrong, it is difficult to pinpoint the problem from the mass of interrelated nodes and links in the network. These problems along with their inability to interpret the output are major disadvantages of these systems [19] [29].

3.3.3.3 Neural networks and fis

In the past few years, NNs have received extensive attention and exploration from the FIs. The reason for this attention is the dynamic and evolving nature of the fraud detection application. Overall, neural networks have shown effective results in areas such as fraud detection by looking at massive quantities of data which have a number of independent variables. These systems have been trained to find patterns and correlation among the incoming transactions.

3.3.3.4 Fds and credit card fraud

As discussed, FIs make extensive use of NN based software to spot and flag transactions inconsistent with the cardholder's usual behavior. The focus of attention in this research is FDS, a NN base software being used by 40 of the top 50 [22] large credit card issuers worldwide including our collaborating FI. Historically, the first version of this software entered the market in 1992 [27]. FDS is a real-time customized software designed to determine the likelihood of card fraud. By using legitimate and fraudulent transactions, FDS has built an individual behavior profile for each account. To the knowledge of the author, there is no documentation on the software, due to the proprietary and business concerns of the software provider. Therefore, it is not clear how this profile is established but the conjecture is that the account profile file includes the type of merchant at which the cardholder typically shops, the time of the day that the cardholder normally makes purchases, the geographic locations along with many more characteristics that only software developers are aware of. FDS inspects and evaluates the incoming transactions to see if they fit into the customer's established profile. Any deviation from the usual cardholder's behavior is monitored and scored by this system. Based on the changes that FDS detects in the customer's pattern of behavior, it assigns scores between I and 1000 to each transaction. The higher the score, the higher the likelihood of fraud. Bank authorities set a threshold value and all transactions scored above this threshold are considered suspicious so that when these scores hit the set threshold, a case is created and is flagged for further investigation. Inherently FDS makes no assumption about the suspicious transactions and transmits the flagged accounts, in real time, to the FI's fraud department for further follow up and investigation [12].

3.4 FRAUD INVESTIGATION PROCESS

To prevent more losses due to credit card fraud, FIs have set up groups or departments responsible for following up on the potentially suspicious transactions identified by the FDS. The flagged accounts with their associated transactions are presented on the fraud analyst computer screen in real time. Fraud analysts examine the flagged transactions with the client's history and from their experience determine the potential risk associated with these transactions. This judgement is based on different criteria such as the type of the merchandise (e.g., jewelry, high price electronic items), the unusual number of transactions or large amount of charges in a given day, the credit limit variations. Based on bank policy, whether a transaction is considered to be legitimate or fraudulent, the fraud analyst has to call the cardholder for transaction verification. In general, in a fraud investigation process, the following possibilities might occur:

- The cardholder can be reached- The cardholder confirms the transaction, referred to as 'false positive'. Approximately 90 percent of flagged cases by FDS are false positives. The cardholder denies the transaction which results in two possibilities:
  1- The card is lost, stolen or counterfeit,
  2- The cardholder has made the purchase but repudiates the event by reporting the card as lost or stolen. In both cases the fraud analyst will block the account.
- The cardholder cannot be reached- The investigator will leave the customer a message to call the bank back as soon as possible, s/he may block the account temporarily and makes a note on the system for further follow up.- The analyst is not able to find the cardholder due to wrong address or telephone number. This case has the high potential of fraud; therefore, the account will be blocked. This procedure will be repeated for all flagged accounts.

3.5 FRAUD DETECTION DILEMMA

Credit card fraud detection is a pattern recognition problem. Every cardholder has a shopping behavior which establishes a profile for her/him. As the result of personal needs or seasonal reasons, patterns of behavior change over time so that s/he may develop new patterns of behavior, which are not known as yet by the Fraud Detection System (FDS). Very often an 'unusual* transaction is legitimate. It is notable that the terms legitimate and non-fraud are equivalent and throughout this thesis are used interchangeably. Currently, FDS identifies

many legitimate accounts as fraudulent resulting in a large number of false positives (FPs). As every cardholder has a huge number of possibilities for developing new patterns of behavior, the types of transactions are widely variable. Hence, it is almost impossible to identify consistent and stable patterns for all the transactions. In fact, there are so many variations of behavior for each individual that are exponential in combination and the complexity of enumerating all combinations of cases are enormous. This ever- changing pattern of behavior along with the combination of legitimate and fraudulent cases has left the FIs with a large number of FPs (approximately 90% of flagged accounts) that has to be investigated. The motivation of this research is to address these challenges. In brief, the task is to postprocess the FDS output and to identify the legitimate transactions (True Negatives, TN) from the stream of flagged transactions. This identification is a classification task, that is, the system we develop has to be able to extract the True Negatives (TNs) from the pool of data while not missing fraudulent transactions. If this goal could be achieved then the bank staff may not need to call these legitimate customers for transaction verification. Pattern recognition for these occurrences is inherently complex and one has to understand the underlying system as much as possible and use this knowledge in the design of the required system. Investigation of some of the AI methodologies and their application revealed that learning is the appropriate approach for addressing this type of classification problems. In fact, learning is very much appropriate for cases where patterns of behavior in real world problems are complex and there is little or no knowledge of the semantics of the application domain. A further survey on some of the learning methodologies and their application led to learning decision trees methodology for this research topic.

3.6 SUMMARY

This Chapter introduces the existing fraud solution approaches and gives a brief introduction to the existing Fraud Detection Systems (FDS). It touches on neural network technology, its advantages and disadvantages and briefly describes its application to credit card fraud detection. It further elaborates on the fraud investigation process and the associated. Issues.

4. METHODOLOGY

4.1 OVERVIEW OF LEARNING SYSTEMS

A learning system is a computer program that makes decisions based on the accumulated information contained in the available known samples. A typical learning system is designed to work with some general model, such as a decision tree, a discriminant function, or a neural net. Different learning systems use a variety of techniques to extract the knowledge from the learning set. These techniques include many highly mathematical methods that can search systematically over large numbers of possibilities to find the closest fit to the data [30].

4.1.1 The Classification Model

To train and evaluate a learning system, the available data should be divided into three parts:

1. The training set.
2. The testing set.
3. the case set.

The training set is used to extract the maximum amount of information from the samples. The testing set is used to estimate the accuracy of the trained system and is a stage where the trained system is validated. The case set is used to evaluate the prediction accuracy of the classifier on future cases. Although extensive computer processing is required by any learning system, For a given problem, atl the very least s/he must describe and define the relevant set of observations and objectives. All the observations are symbols that are being manipulated by the computer. Thus,while the computer can carryout different forms of analysis, of the potential for success lies with the analyst who select the real world data with the required accuracy [30].

4.1.2 Hypothesis Space in Supervised Learning

In supervised learning, the training examples of the form $\{(a_1, b_1),.,(A_m > B_k)$ for some unknown function $y = f(x)$. The x_i typically represents discrete or real valued components such as color, height, or age with their associated values. The y values typically represent a discrete set of classes { l,..., k}. The task of learning program is: given a set of training examples of f, return a function them approximates. This function is a hypothesis about the true function Any

preference for one hypothesis over another, beyond more consistency with the example is called bias. Because usually there are a large number of possible hypotheses A simple example consist shown in Figure 4.1 is used to clarify the meaning of hypothesis and bias in this context. shows , (x, y) points represents the training examples, where y = f(x). The task of the learning algorithm is to find a function (x) that fits these points as closely as possible. As Figure 5.1 shows, the function used in (b) is a piecewise linear function, in (c) it is a more complicated function (e.g., quadratic) and in (d) a Least Square function is used to fit to the data points. As discussed above the true f is unknown and different functions of h. try to approximate f by finding a function that is a good fit to the available data samples. Any preference of (b) over (c), (d) or any other possibility is considered a bias.

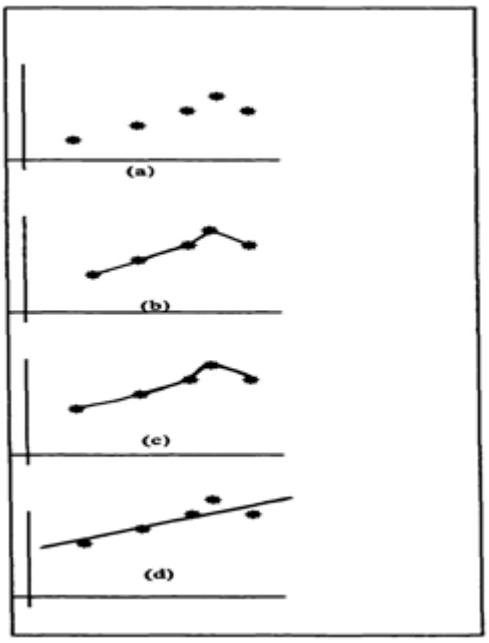

Figure 4.1: illustration of Several Hypothesis in Learning Algorithm [RUSS95].

4.2 PERSPECTIVES ON CLASSIFICATION

Historically the strands of research on classification can be represented in three main and distinct categories: (1) statistical, (2) neural networks (NNs), and (3) machine learning (ML). As explain before, the goal of classification is to derive rules or procedures that would be able [19]:

- To equal, if not exceed, a human decision maker's behavior, but have the advantage of consistency.
- To handle a wide variety of problems and given enough data, could be generalized.

For this purpose there are differen algorithms that search a hypothesis space defined by some un derlying representation (e.g., linear function, neural networks, logical descriptions, or decision tree) For each of these hypotheses representations, the corresponding learning.algorithm takes a dvantage of a different underlying structure to organize the search through the hypothesis space .Statistical methods are considered parametric, whereas NN and ML methods are categorized as nonparametric. Parametric methods assume a certain form of the underlying model, such as a n ormal (bellshapd) curve for the classifier. Non parametric methods make no assumption about the functional form of the underlying model. These methods employ the power of computers to search and iterate until they find a good fit to the sample data [19].

4.2.1 Neural Networks

Neural networks consist of layers of interconnected nodes, each node producing a non linear function of its input. The input to a node may come from other nodes or directly from the i nput data Some nodes are also identified with the output of the network. The complete network, t herefore, represents a very complex set of interdependencies, which may incorporat different deg rees of nonlinearity, allowing very general functions to be modeled [13].

4.2.2 Machine Learning

Machine learning is inherently a field. It draws on results from artificial intelligence (AI), probab ility and statistics, computational complexity theory, control theory, information theory, philos ophy, psychology, obtained from a training set of preclassified cases, are used to predict the clas ses of new cases [19].

Machine learning methods have been applied to a variety of large databases to leam general reg ularities implicit in the data. For instance, decision tree learning algorithms have been used by N ASA to leam how to classify celestial objects from the second Palomar Observatory Sky Survey . This system is now being used to automatically classify all objects in the Sky Survey, which co nsist of there terabytes of image data [20].

4.3 LEARNING DECISION TREES

Decision trees,a machine learning method,are perhaps the oldest,and one of the most popular wa y to represent the outcome of classification learning procedure. It is a method for approximating discrete-valued target functions, in which the learned function is represented by a decision tree [11].Decision trees are capable of representing the most complex problems given sufficient data , and they are one of the most highly developed techniques for partitioning samples into a set of decision rules. Learned trees can also be represented as sets of ifthen rules to improve the human readability. These learning methods are very popular and have been successfully applied to a broad range of tasks from learning to diagnose medical cases to learning to asses credit risk of loan applicants [20].

4.3.1 Domain Application of Decision Tree Learning

Although a variety of decision tree learning algorithms have been developed with somewhatdi fferent capabilities and requirements, decision tree learning is generally best suited to problem s with the following characteristics [20]:

- The target function has discrete output values.For instance,decision tree assigns a ‘yes’ or ‘no’ to each classified example.
- The training data may contain errors.Decision tree learning methods are robust to error s found in the attribute values that describe the input examples.
- The training data may contain missing attribute values. Even though the value of some of the training examples might be unknown, still decision tree learning methods can be employed.

It was realized that due to these characteristics decision tree learning is a suitable fit for this research topic.

4.3.2 An Illustration of Decision Tree Induction

To visualize how a decision tree learning algorithm leams from the training set, the following example has been adapted from Russel [24]. The task is whether to wait for a table at a restaurant or not. The aim is to leam a decision for the concept WillWait by employing decision tree methodology. As the first step, the features that can describe the examples are as follows:

1. Alternate: whether there is a suitable alternative restaurant nearby.
2. Ban whether the restaurant has a comfortable bar area to wait in.
3. Fri/Sat: true on Fridays and Saturdays.
4. Hungry: whether we are hungry.
5. Patrons: how many people are in the restaurant (values are None, Some, and Full).
6. Raining: whether it is raining outside.
7. Reservation: whether we made a reservation.
8. Type: the kind of restaurant (French, Italian, Thai, or Burger).
9. WaitEstimated: the wait estimated by the host (0-10,10-30,30-60, >60 minutes).

The decision tree that can represent this task is shown in Figure 4.2. The tree can be describedas a conjunction of individual implications corresponding to the paths through the tree ending into Yes/No nodes. As an example, the path for a restaurant full of patrons with an estimated wait of 10to30 minutes when the person is not hungry can be expressed by the following logical sentence:

∀ r Patron (r, Full) Λ WaitEstimate (r, 10-30) Λ Hungry (r, N) => WillWait (r)

The notation employed is defined below:

r: a general indictor for representing a person or object

N: means 'No'

∀: for every person or object

Λ : and

=>: then

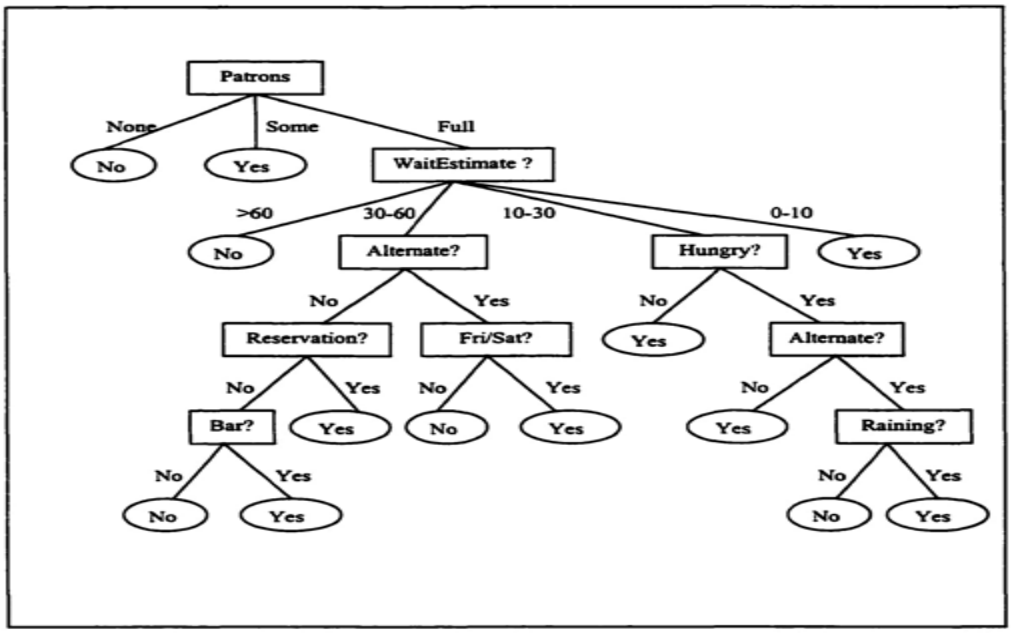

Figure 4.2: A decision Tree for Deciding Whether to Wait for a Table in a restaurant [RUSS95].

4.3.2.1 Induction of decision trees from examples

ln this section the sets of 12 examples (X_1,… …… X_{12}) along with their value features, and the value of the class associated to these features are illustrated in Table 4.1 It should be noted that when the goal is true for some examples they are called positive examples and when it is not true they are called negative examples. As Table 4.1 shows the positive examples are the ones that have the value of Yes (e.g., X_1, X_3,..) for the goal WillWait and the negative examples are the on es that have the value of No (e.g., X_2, X_5,..)for this goal.The complete set of 12 examples is calle d the training set.The task is to find a decision tree that agrees with all the examples. A trivial sol ution to this problem is to construct a decision tree that has one path to a leaf for each case where the path tests each feature in turn and follows the value for the example, and the ending leaf has the classification of the example. If this route is taken for learning a decision tree, with the occur rences of the same examples, the decision tree will obviously come up with the right classificati on without any errors. But this tree is not able to classify other cases correctly because this trivia l tree has just memorized the observations and has not extracted any pattern from the examples. If a learning algorithm does not extract general rules from the data it will not be able to extrapola te to new cases. That is why the learning algorithm looks at the examples, not at the correct func

tion.While pondering this simple example, one can understand why the learning algorithm has e rrorsin the process of training even though the true class of the examples is presented to it.

Table 4.1: A Small Training Set for the restaurant Domain.

Examples	Features									Goal WillWait
	Alt	Bar	Fri	Hun	Patron	Rain	Reserve	Type	Estimate	
X1	Yes	No	No	Yes	Some	No	Yes	French	0-10	Yes
X2	Yes	No	No	Yes	Full	No	No	Thai	30-60	No
X3	No	Yes	No	No	Some	No	No	Burger	0-10	Yes
X4	Yes	No	Yes	Yes	Full	No	No	Thai	10-30	Yes
X5	Yes	No	Yes	No	Full	No	Yes	French	>60	No
X6	No	Yes	No	Yes	Some	Yes	Yes	Italian	0-10	Yes
X7	No	Yes	No	No	None	Yes	No	Burger	0-10	No
X8	No	No	No	Yes	Some	Yes	Yes	Thai	0-10	Yes
X9	No	Yes	Yes	No	Full	Yes	No	Burger	>60	No
X10	Yes	Yes	Yes	Yes	Full	No	Yes	Italian	10-30	No
X11	No	No	No	No	None	No	No	Thai	0-10	No
X12	Yes	Yes	Yes	Yes	Full	No	No	Burger	30-60	Yes

To find a pattern from the examples means to find some regularities in the training set and to the be able to describe alargenumber of examples in a conciseway.The whole point of the decision tree is to find ways that only parts of the inpu needtobe incorporated in the structur of the tree to reach a decision.Inother wordsthe decision tree algorithm tries to find a small tree thatcorrectly classifiesmost of the training examples.This is an example of ageneral of inductive learning oft en called Occam's Razor: " The most likely hypothesis is the simplest.one that is consistentwith all obsens. "The basic idea behind decision tree learningalgorithm is to test the most important valuefirst.The most important is the one that makes the most difference to the classification of a example. This approach may lead to the correct classification with a small number of tests, mea ning that all paths in the tree will be short and the tree as a whole will be small [24].Figure 4.3.ill ustrates how a simplified version of the algorithm starts. In the first stepThe12 training.example s are classified into positive and negative sets. Then the algorithm starts by deciding which attri bute to use as the first test in the tree. It considers all possible attributes in this way and chooses t he most important one as the root test. As Figure 3(a) shows, Patrons is a fairly important attribu te because if its values are None or Some, then it leads to example sets for which the classificati on is definitely No or Yes, respectively. If the value of this test is Full then further tests are requi

red. As Figure 3(b) shows Type is a poor attribute, because it has four possible outcomes each of which has the same number of positive and negative values All possible attributes are considere d in this way and the most important one is selected for the root of the tree. Supposing that in thi s example the most important attribute is Patrons, it is considered as the root test. After Patrons splits up the examples, each outcome is a new decision tree learning problem in itself, with the f ewer examples and one fewer attribute (Patrons has already been picked up).

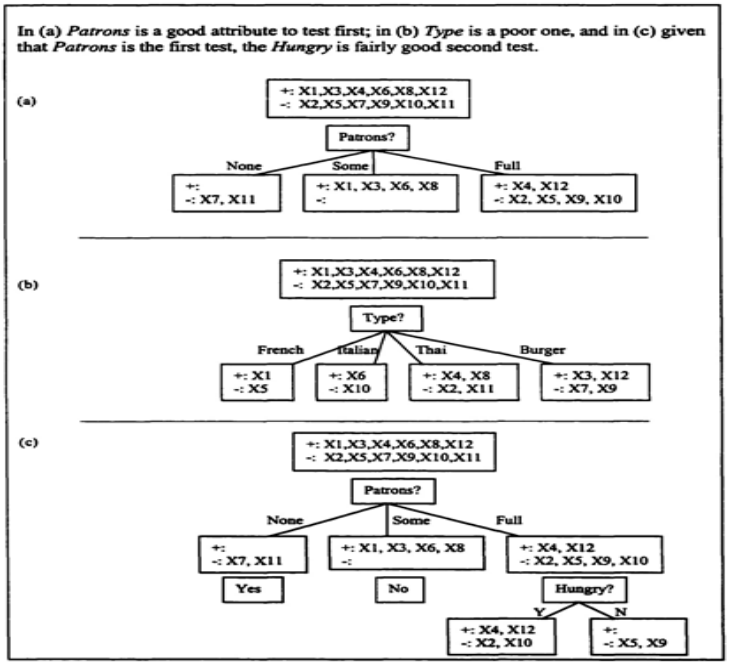

Figure 4.3: Partitioning the Examples by testing on Attributes [RUSS95].

Figure 4.4 illustrates this partitioning. This Figure shows that decision tree learning algorithms can be seen as a method for partitioning the universe into successively smaller rectangles with the goal that each rectangle only contains objects of one class, that is, positive or negative [19]. The dashed line shows the real division of examples in the universe. The solid lines show a decision tree approximation.

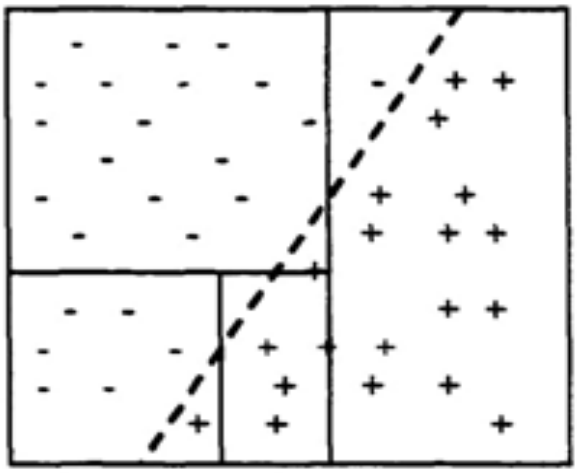

Figure 4.4: Decision Tree Learning Algorithms Partitions The Universe Into Successively Smaller Rectangles [MICH94].

In general, three possibilities can be considered for decision tree problems [24]:

1. If there are some positive and negative examples, choose the best attribute to split them. In Figure 4-3 (c) this fact is illustrated by using Hungry to split the remaining examples.
2. If all the remaining examples are positive (or all negative) then the search is over.
   For instance in Figure 4-3 (c), the answers of yes or no are assigned to None and Some, respectively.
3. If there are no examples left, the algorithm returns a default value, which is calculated from the majority classification at the node's parent.

The decision tree learning algorithm applied to this problem is shown in Figure 4.5 algorithm continues until the tree shown in Figure 4.6 is constructed. As it can be seen this tree is distinctly different from the original tree shown inFigure 4.2 despite the fact that the same sample data were used to generate it.One might concludethat the learning algorithm is not learning the correct function.This conclusion is no correct because as ment ionbefore,the learning algorithm looks at the examples,not atthe correctfunction,and in fact, its hypothesisnotonly should agree with all thecases,but should be considerably simpler that the original tree.The learning algorithmhas not considered any test for Raining and Reservation because has been able to classify all the examples without them.The algorithm

also has detectedan interesting regularity in the data,that is, the person will wait for Thai food on weekends [24].

```
Function Decision-Tree-Learning (examples, attributes, default) returns a decision tree
    Inputs: examples, set of examples
            attributes, set of attributes
            default, default value for the goal

    if examples is empty then return default
    else if all examples have the same classification then return the classification
    else if attribute is empty then return Majority-value (examples)
    else
        best ← Choose-Attribute (attributes, examples)
        Tree ← a new decision tree with root test best
        for each value v_i of best do
            examples_i ← {elements of examples with best = v_i }
            subtree ← Decision-Tree-Learning (examples_i , attributes – best,
                                              Majority-Value (example)
            Add a branch to tree with label v_i and subtree subtree
        end
        return tree
```

Figure 4.5: The Decision Tree Learning Algorithm [RUSS95].

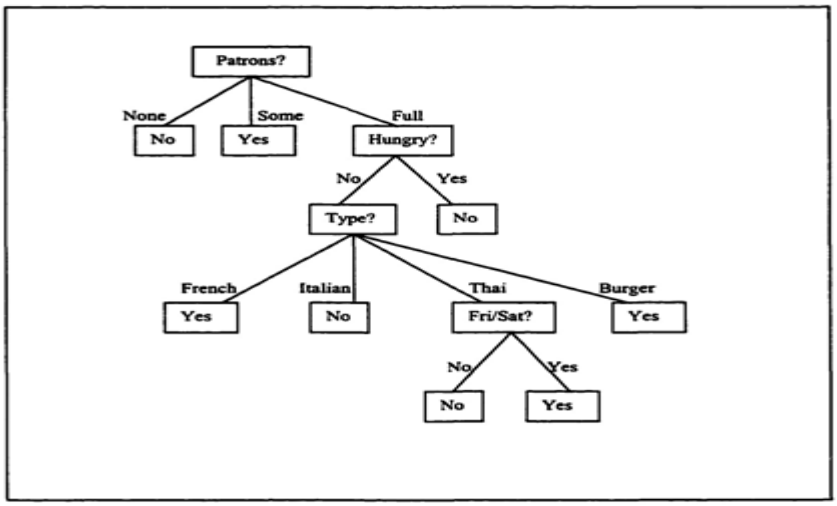

Figure 4.6: the resulting Decision from the 12 training examples.

4.3.3 Boosting

Boosting is a techniquefor generating and combining multiple classifiers, eitherdecisiontrees or rulesetsThis technique is used toimprove the prediction accuracyof the classifiers.Boosting may lead to a reduction in error rate but this effect is not guaranteed and in some cases it might have no effect at all.The effectiveness of boosting is not deterministic and it is not known beforehand.Only afteremploying this technique on the data and comparing the results one can see whether theprediction accuracy has improved or not. In boosting, instead of one classifier, several classifiers are constructedand the combination of their outcomes will determine the final class being assigned tothe case. Boosting may give higher predictive accuracy at the expense of increased classifierconstruction time [10].

4.3.4 Cross-Validation

One of the techniques for getting more reliable estimates of the predictive accura cy of the classifiers is fold cross.validation (CV). The basic idea of the cross.validation.techn ique.is to.try,to estimate how well the current system will predict the unseen data. The idea is, instead ofusing one sample to build a tree and another sample to test the tree,the algorithm will form several pseudoindependent samples from the original samples and usethese samples to form a more accurate estimate of the error. For this purpose,the program splits the data into a number of folds (splits) equal toa chosen number. Experience on a large number of datasethasshown that the number of folds equal.to 1 0 has achieved go od results. That is why in many learning algorithms the number of folds is chosen at10asthe default option.For eachfold in turn, aclassifier isconstructed from the examples in allthe other folds and thenits.accuracy is tested on theexamples in the hold-out fold. In this way, each case is used just once as a test set. The error rate of a classifier produced from all the samples is estimated as the ratio of the total number of errors on the hold-out cases to the total number of cases [19].

4.4 SUMMARY

The goal of a learning system is to extract decision rules from the sample data. Machine learning addresses the problem of how to build computer programs that improve their performance at some task through experience. Major points of this Section include:

- Introduction of the classification notion and main strands of research in this area along with an overview of learning systems and their requirements.
- Designing a machine learning approach involves a number of design choices, including choosing the type of training.representation for this target function, and an algorithm for learning the target function from the training examples.
- Learning involves searching through a space of possible hypotheses to find the hypothesis that best fits the available training examples.

Description of learning decision trees and its domain of application.

5. APPLICATION OF CREDIT CARD

5.1 DATA REQUIREMENTS

The first step is to choose the type of training examples from which the system will learn. The learning algorithm will extract the patterns of behavior from the examples fed into it, therefore, the whole application is dependent on the information contained in the data set. In general, data collection for a real world system will have limitations, that is, there will be some information lost or not provided and we are restricted to what is available. The operation of the FI transaction authorization and tracking system was described in Chapter 2. As Figure 2.1 showed the transaction tracking system of the FI communicates with FDS in real time and transmits all the incoming authorizations from the POS to this system to be scored. When the transactions get to a point where their FDS score hits the current threshold, a case is created and passed on to the fraud department for further investigation. In this way FDS detects the suspicious transactions from the stream of transactions. This output includes both the real fraudulent transaction (True Positives, TP) hit by the system and false positives (FPs). It is important to note that there are fraudulent transactions that are being missed by the FDS; because their score is below the threshold and, therefore, they are being missed. This category of transactions is known as False negative (FN) which means that the case was fraud but the system missed it. As discussed, FDS has already identified some, perhaps most, of the TPs but in the meantime it has created a lot of FPs as well. The task remains to us is to process these cases and to identify as many real legitimate transactions as possible while trying not to miss fraudulent cases (FNs).

5.2 DATA COLLECTION

To perform the analysis, the accounts flagged by FDS were used as the input of the learning system. The data was provided by the collaborating FI. The transactions flagged by FDS are taken over 45 days (June, July, and part of August 99) and are related to a limited region of Toronto. Together, ten separate files were provided. The first nine files were related to flagged confirmed legitimate accounts, which together consisted of 4919 accounts with 69,182 transactions. Due to the volume of data for the legitimate accounts, they were divided

into nine separate files. The tenth file included 707 fraudulent accounts that contained 6,725 transactions. It should be noted that the fraudulent accounts have a combination of fraud/non-fraud transactions. Hence, the fraudulent accounts consisted of1,743 legitimate and 4,982 fraudulent transactions. Due to the confidentiality of real account numbers and in order to have all the transactions from each account together, a substitute but unique number was assigned to the transactions of each account by the FI. A very small sample of the data in raw format is available in Appendix A. All ten files had the same fields and each transaction had the following information:

- A replacement account number
- Date/Time of transaction
- Transaction amount
- Merchant country code
- Merchant category code (SIC)
- Decision code
- POS
- Type of card
- Case creation
- First action

Although the scores associated to each transaction by the FDS, were of great importance for the analysis by the learning system, due to the proprietary and business concerns of the software provider, the FI was not able to provide this information. In the meantime it was essential to identify which transactions deviated from the normal behavioral pattern of the legitimate cardholders which caused the system to flag them as potentially fraudulent. The FDS scores could show this trend, none the less to make up for this data shortcoming, the case creation date was provided as a proxy to each transaction. Lack of scores not only may have serious impact on the precision of the classifier, but also due to the high volume of data, it caused uncertainty and substantial amount of ‘manual’ work in selecting the transactions that occurred close to case creation date.

5.2.1 Labeling the Transactions

To use the data for training, it was necessary to identify the fraudulent transactions from the legitimate ones. Currently, labeling the fraudulent transactions is done manually and the fraud investigation department keeps conventional paper based fraud files on which they mark the transactions that were identified as fraud. Due to this manual process, there is no mechanism to migrate this information back into the transaction tracking system and, therefore, there is no record keeping of them on the system. Fraudulent accounts normally have a mixture of fraudulent and legitimate transactions, therefore, the confirmed fraud transactions in fraud file were labeled by the bank with an asterisk (*).

5.2.2 Preprocessing the Databases

Raw data contains the information that must be extracted but in the meantime it contains too much non-essential information. The raw data provided by the FI, required substantial preprocessing to weed out the irrelevant information and to prepare the data set in a suitable form for the learning system. The original data files were in text format, therefore, Excel was selected as a tool for data manipulation.

The first step was to go through all the transactions and find the closest set of transactions which match the case creation date. In the absence of scores, this can partially help to identify those chains of transactions which, from the FDS point of view, did not have normal behavior and caused the system to score them gradually and eventually get to a point where they hit the threshold set by the bank. Manual inspection of data revealed the existence of some inconsistencies in the data. To be able to use the data for the analysis these inconsistencies had to be removed from the datasets. After the preprocessing of databases, the final legitimate database consisted of 13,426 non-fraud transactions and the final fraudulent database consisted of6,666 transactions (4,969 fraud and 1,698 non-fraud).

After the initial set up of the databases, the integrity of the data had to be investigated. The data had fields with unknown values, spaces or zero values. All unknown values were replaced by a question mark “?”. Letters such as A, D, R, P, K, S, Y, N, were checked to be in one format, that is, in capital letters. Meanwhile, all non-fraud transactions in both databases were labeled with the letter “N”.

5.3 LEARNING REQUIREMENTS

Learning means behaving better as the result of experience. The task of a learning system is to extract the maximum amount of information from the data samples, and based on this information, to estimate the accuracy of its future classifications and predictions. While conceptually simple, extracting information from a large database requires careful organization and the specification of the goals to be met by the learning system. The simple requirement of the classification methods is that the data be presented in the form of samples composed of patterns of observations with the correct classification. Then the learning procedure will be applied which is an iterative process [30].

5.3.1 Features and Classes

In the problem of predicting whether a flagged transaction is fraud or non-fraud, there are two classes: fraud and non-fraud. The task is to predict which is the correct class based on the observations of a set of transactions. By employing a decision tree learning algorithm, the aim is to learn a definition for the concept, Transaction (fraud/non-fraud), where the definition is expressed as a decision tree. In setting this up as a learning problem, the properties or features that are available to describe the examples are presented in Table 5.1. A small sample data set for credit card transactions obtained from the FI is available in Appendix A.

Table 5.1: Credit Card Observation.

Feature	Description
Account No.	Cardholder's account number
Date/Time	Date and time of transaction
Dollar	Dollar amount of transaction
SIC	Merchant category code
Country	Merchant country code
Decision	Authorized (A), Declined (D), Referral (R), Pick up (P)
POS	Card swiped (S) / keyed (K)
Card type	Classic / Gold
Case creation	The day / time case created by FDS
Case action	The day / time fraud analyst started to investigate on the case

5.4 CONCEPT LEARNING AND SEARCH SPACE

The problem of finding general functions from specific training examples is central to learning .Concept learning is acquiring the definition of a general category.given asample of positive and negatives training examples of the category.Concept to learning can beviewed as thetask of searching through a large space of hypotheses,implicitly defined by the hypothesis representation (e.g., decision trees), to find the hypothesis that best fits the trainin gexamples [20].In general, a well-defined learning problem requires a well-specified task, source of training experience, and performance metric [20]. Applying these criteria to this research application results in the following descriptions:

- Task T: Transaction is either fraud or non-fraud
- Training experience: A database of legitimate and fraudulent transactions with their labels.
- Performance measure: Percentage of cases classified correctly by the classifiers.

To complete the design of the learning system, the following factors should be chosen:

- The exact type of knowledge to be learned (i.e., classifying the transactions as fraud/non-fraud).
- A representation for this target knowledge (i.e., decision trees)
- A learning mechanism (i.e., a learning algorithm)

5.4.1 Software Selection

If learning is viewed as a search problem,then it is natural that learning algorithms will exam different strategies for searching the hypothesis space. The algorithms that are capable of Efficiently searching very large hypothesis spaces to find the hypotheses that bestfit training data are of great interest.In the field of ML,a variety of programs have been developed.Three of these software are: C4.5[QUIN93|,CART [2], and RIPPER (COHE951.C4.5and CARTare decision tree learning based softwarewhereas RIPPER is rule based learning system.RIPPER was not chosen because it is a Unix based system and there was no access to Unix systems.T he PC version of CART is available but it was costlier than C4.S. Moreover, CART can only produce decision tress whereas C4.S is able to transform the enerated trees into a set of rulesC4.S is a Unix based system.The PCversion of this software is also available called See.

5.4.1.1 Trees into Rules

There has always been an argument in favor of rulebased representations over tree structure representations,on the grounds of readability and userfriendlinessWhen domain is complex, decision trees can become very “bushy” and difficult to understand, whereas rules tend to be

modular and easier to understand.On the other hand, decision tree construction programs are usually very fast.A compromise is to use a decision tree algorithm to build an initial tree and then derive rules from the tree, the trees into a set of rules [22].This functionality is impleme nted in See5.

Moreover, a rule set generated from a tree usually has fewer rules than the tree has leaves. A simple example adapted from [11] shows the reason for this compactness.Consider the propositions (A=l and B=l) or (C=l and D=l). If each of the four attributes of A, B, C, and D, has two possible values (e.g., 1 and 0), the proposition represented by the rule sets is as follows:

Rule I: A = 1 and B = 1 -> +

Rule 2: C = 1 and D = 1 -> +

Rule 3: otherwise -► -

If feature A is arbitrarily selected as the partitioning criterion for the root node,

The most compact single decision tree representation for this rule set is shown in Figure 5.1. Obviously this decision tree is less understandable than the above rule set. This readability Problem corresponds to the number of possible paths through the tree. Just Rules 1 and 2 generate three paths in the tree, shown in Figure 5-1:

A = I and B = 1 —> +

A = 1 and B = 0 and C = 1 and D = 1 —▶ +

A = 0 and C = 1 and D = 1 -> +

In general,many functions with small prepositionalor rule representations have corresponding decision trees that are large, redundant, and inefficient [11]. For very large data set, however, generating rules can require considerably more computer time than generating the trees.

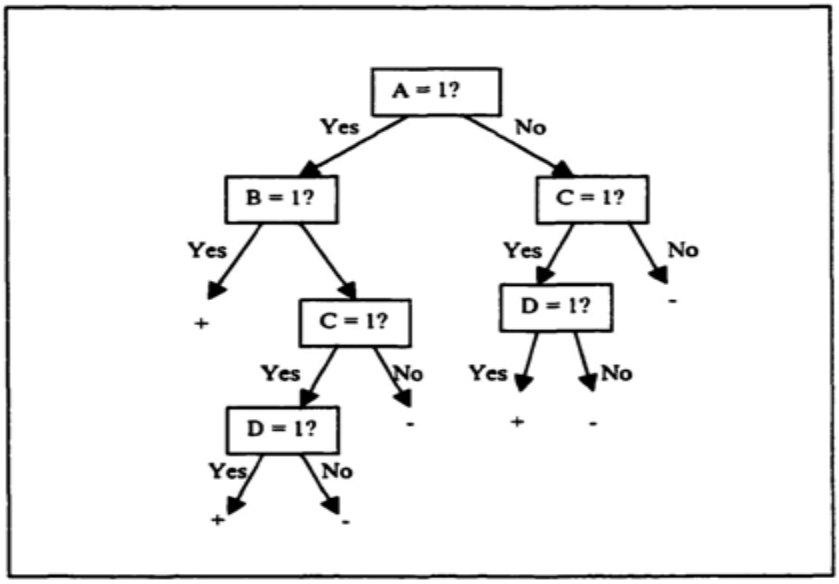

Figure 5.1: Decision Tree Representation.

5.5 SEE5

See5 is adecision tree learning software package which was designed and developed by Ross Quinlan, an scholar, pioneer,and researcher in the field of machine learning for many years. Quinlan is the director of the Rulequest Research Institute,located in Australiaand.See5could be purchased from him through the Internet See5 is a learning system that extracts informative

patterns from the data.It analyzes the data to produce decision trees and/or rulesets that relate

a case's class to the value of its features. The following sections introduce the See5 options that were employed in the analysis for the current application.

5.5.1 See5 Construction Options

As the default option, See5 constructs a decision tree. A set of training and testing data was selected and the program ran with See5 default option. The result is shown in Figure S-2. Although the program gives all the details of the individual decision trees, due to their large size, only the output summary of the learning sets is shown in this Figure. The first section shows the evaluation results of the decision tree, first on the training set from which the tree wasconstructed,and.then.on.the.test.set.The size of the tree shows the number of leaves and the column headed errors, represents the number and percentage of the cases misclassified by the tree. The tree, with 193 leaves, misclassifies 1,816 of the 13,405 cases, thus having an error rate of 13.5%. Performance on these.cases is further analyzed in a confusion matrix t hat pinpoints the kinds of errors made. In this example, the decision tree misclassifies 370 (3.6%) of the legiti mate cases as fraudulent and 1,44 (42.5%) fraudulent cases as legitimate.For the test set, the tree with 193 leaves, misclassifies 929 of the 6,465 cases, thus having an error rate of 14.4%. The confusion matrix for the test set again shows the detailed breakdown of correct and incorrect classifications. The decision tree misclassifies 545 (10.7%) of the legitimate cases as fraudulent and 384 (27.2%) fraudulent cases as legitimate.

One might ask why the training algorithm makes any errors in the training phase while it is classifying the cases where the outcome is known. One should keep in mind that the essence of learning is to move beyond the training samples. Thus, the learning algorithm does not memorize the cases it has seen but rather its attention is extracting rules and patterns of behavior from the data to be able to generalize and extrapolate them to future cases.

```
Evaluation on training data (13405 cases):

          Decision Tree
        ----------------
        Size      Errors

         193  1816(13.5%)   <<


         (a)   (b)    <-classified as
        ----  ----
        9679   370    (a): class N
        1446  1910    (b): class Y


Evaluation on test data (6465 cases):

          Decision Tree
        ----------------
        Size      Errors

         193   929(14.4%)   <<


         (a)   (b)    <-classified as
        ----  ----
        4511   545    (a): class N
         384  1025    (b): class Y
```

Figure 5.2: Output Summary Of the Learning Set.

5.5.2 Rulesets

Decision trees can sometimes be very difficult to understand. An important feature of See5 is its ability to convert trees into collections of rules called rule sets. The same learning set ran by the Rulesets option of See5. Here again, the program gives all the details of the individual decision trees and rule sets but for the sake of brevity only the output summary of the learning sets is shown in Figure 5.3As can be observed, the decision tree with 193 leaves is reduced to 81 rules but the rules have a slightly higher error rate than the trees(0.2%).The rulesets option, with 81 rules, misclassifies 1,840 of the 13,405 cases, thus having an error rate of 13.7%. Performan ce on these cases is further analyzed in a confusion matrix that shows the types of errors made. In this example,the rule misclassifies 291 (2.8%) of the legitimate cases as fraudulent and 1,549 (46.2%)fraudcases as legitimate.For the test set, the rule sets misclassify 877 of the 6,465 given cases, showing an error rate of 13.6%. The confusion matrix for the test cases again shows the d etailed breakdown of correct and incorrect classifications. The rule sets misclassify 481 (9.5%) of the legitimate cases as fraudulent and 396 (28.1%) fraudulent cases as legitimate.

```
Evaluation on training data (13405 cases):

          Decision Tree           Rules
        ----------------     ----------------
        Size      Errors      No      Errors

         193  1816(13.5%)      81  1840(13.7%)    <<

                              (a)    (b)     <-classified as
                              ----   ----
                              9758    291     (a): class N
                              1549   1807     (b): class Y

Evaluation on test data (6465 cases):

          Decision Tree           Rules
        ----------------     ----------------
        Size      Errors      No      Errors

        193    929(14.4%)      81    877(13.6%)    <<

                              (a)    (b)     <-classified as
                              ----   ----
                              4575    481     (a): class N
                               396   1013     (b): class Y
```

Figure 5.3: Output Summary Of The Learning Set.

5.5.2.1 Boosting

Boosting was introduced in Section 4.4.4. All the steps required for the construction of different classifiers in this procedure are embedded and implemented in the learning algorithm, by the software developer, and normally there is no documentation on the details of these procedures due to proprietary issues. The Boost option with 10 trials was selected and the program ran for the same learning set. The summary of the results is shown in Figure 5,4. As the first step, a single decision tree or rule set is constructed as before from the training data. This classifier will usually make mistakes on some cases in the dataset (Trial 0 in Figure 5,4). When the second clas sifier is construct, the algorithm pays more attention to the misclassified cases to try to get them right. This makes the second classifier different from the first one (Trial 1 in Figure 5,4). The second classifier will also make errors on some cases, and these become the focus of attention during the construction of the third classifier (Trial 2). This process continues for a predetermine d number of iterations. The Boost option with x trials allows See5 to construct up to x classifiers in this manner (suggested default is 10). Naturally, constructing multiple classifiers requires extra computational time and resources but the effort might be worth the cost. Different ML sou rces and trials over numerous datasets, large and small, have shown that on average the 10 classifier boosting is the most appropriate choice [QUIN99].

It should be noted that Boosting trials greater than 10 were also examined in the experiments performed, however, it never exceded 10 trails before the algorithm terminated.One example is illustrated in Figure 54.It is interesting to note that although the number of trials for the boost option was set to be 10, the algorithm terminated after 7 trials. This shows that the software can determine when there is no improvement possible on the accuracy reached. The classifier performance is summarized for each trial on a separate line, while the line labeled boost shows the overall results of all the classifiers [QUIN99]. The constructed by Trial 0 is identical to the one produced without the Boost option (See Figure 5,1). Some of the subsequent trees produced when the algorithm was paying more attention to certain cases have quite high overall error rates . When the seven trees were combined by the functions implemented in the algorithm, the final predictions have an error rate of 11.4% on the training examples.

5.6 DESIGN OF EXPERIMENTS

For setting up the experiments several steps were taken.Databases were randomly split into two main groups by the approximate proportions of 2/3 and 1/3 split [M1TC97] [19] for training and testing sets, respectively. The training set is used to design the classifier, and the testing set is used to evaluate the accuracy of the classifier derived. While sufficient test samples are the key to accurate error estimation, adequate training cases in the design of a classifier are also of great importance. Therefore, the non-fraud database with 13,426 transactions was split into two databases of 8,963 and 4,463 transactions. The fraud database split resulted in 4,442 transactions for training and 2,222 cases for testing. It is important to have a rather large test set for system validation. From 2,222 cases, 2000 transactions were arbitrarily used for testing the classifier and the rest (222 cases) were put aside as case set to evaluate the prediction of the classifier on new cases.In domains, such as credit card fraud, the natural class distribution is between 10:90 and 20:80. This means that between 10 to 20 percent of the flagged cases are fraud (minority instances) and the rest are legitimate (majority instances). In other words, the number of fraudulent transactions is much smaller than the legitimate ones.One of the factors that contribute to the success of a learning process is the class distribution in the training set. Using the same algorithm, different training class distributions can produce classifiers of different quality. Very often using the natural class distribution might not yield the most effective classifier. In other words, using the

natural class distribution for training, might cause the learning algorithm to treat the minority class instances as noise or simply produce classifiers that always predict the majority class instances [2].Related works in fraudulent cellular phone calls or credit card fraud that have class distribution between 10:90 and 20:80, show that class distribution for training is very important and can dramatically improve the performance of the classifiers. Extensive experiments have shown that training data with a 50% fraud distribution produced the best classifiers [6] [28] [15].If the assumption is that the distribution of examples influences the performance of the resulting classifiers, then how can one characterize the effects of the training class distribution on the performance of the classifiers and select a class distribution that can produce the most predictive classifiers? Based on the aforementioned results, the best approach was to create data subsets with different class distributions, then apply the learning algorithm to these subsets and evaluate the effect of class distributions on training by evaluating the performance of the resulting classifiers on the test sets and future cases.

5.7 SUMMARY

This Chapter provides a detailed explanation on data acquirement and the steps required for processing the data to make it suitable for the analysis. It introduces the learning software used and its capabilities. It further elaborates on the data set and class distribution designs for training and testing sets, and the variations of experiments performed.

6. ANALYSIS OF RESULTS AND DISCUSSION

This section presents the experimental results of processing different sets of data with various training class distributions. Unfortunately there was no information available on the costs associated with fraud offices investigations, therefore, potential savings of the methodology developed in this study, could not be estimated.

6.1 STRUCTURING THE RESULTS

Fifty four experiments were performed to study the effects of class distributions on training and variations of different features on the evaluation of the classifiers constructed. The experiments were conducted using See5 construction options of: (1) decision trees, (2) rulesets, (3) boosting, and (4) ten fold cross validation (CV).For each option, the program was run to explore the effects of various class distributions and features on training and testing sets. Although the cross validation technique is typically used for intermediate sample sizes (of order 2000) [QUIN99], however, a decision was made to examine this option in the experiments conducted, to have extra evaluation on the training and testing sets as well.The experiments produced 54 sets of results. Thirty six of these results are the classifiers and the other 18 results present the evaluation of ten fold CV trials on training and testing sets. A selection of output summaries of See5 for several variations of features and class distribution are presented in Appendix B.Tables C-l to C-9 of Appendix C presents the evaluation results of the 36 classifiers on training and testing data sets. Tables C-l to C-3 are related to the training class distribution of 25:75 while exploring the effects of different features in the classifier construction. Tables C-4 to C-6 and C-7 to C-9 are related to the training class distribution of 33:67 and 50:50, respectively. Each Table includes the construction option, the size of the generated trees and / or rules, along with the number of errors and their percentage for both training and testing sets.

6.2 PERFORMANCE ANALYSIS

To do the analysis the classifier attained the lowest error rate among the 36 classifiers presented in Tables C-l to C-9, was selected. As these Tables shows, the boosted decision trees (BDT) classifier trained on 25:75 class distribution attained the lowest error rate of 11.4%. This classifier was selected as the first choice. To compare the performance of this classifier against another one, the decision trees (DT) classifier trained on 25:75 class distribution by attaining the error rate of 13.5% was considered as the second choice.As the fraud rate increases in the training sets, the error rate also increases leading to the conjecture that there is no need for further analysis on the class distribution with higher fraud rates and the above discussed classifiers are the most effective classifiers for further analysis. However, based on the work of other researchers [28] [2] [15] who have employed various training class distribution in their analysis for fraud applications, a decision was made to study the above discussed BDT and DT classifiers not only for 25:75 class distribution but also for the class distributions of 33:67 and 50:50.As discussed, in situations where different types of errors have different costs such as in credit card fraud detection, the elements of the confusion matrix such as TN, FP, FN, and TP are the essential metrics for the system's performance. Therefore, these metrics were considered to be the true indicators for the performance evaluation of the selected classifiers.To compare these metrics for the selected BDT and DT classifiers, a new set of Tables were formed. Tables 6-1 to 6-6 and 6-7 to 6-12 are related to the evaluation of the selected BDT and DT on the training and testing sets. Each Table depicts error rate, TN, FN, TP, FP and their associated rates for each class distribution and the sets of features. TN, TP, FN, and FP rates were calculated based on the information available from the confusion matrix of each classifier, some of them, reported in Appendix B.To visualize the performance of these classifiers, two plots (Figures 6-1 and 6-2) were prepared. Figures 6-1 was plotted based on information obtained from Tables 6-1 to 6-3, and 6-7 to 6-9 illustrating the performance of the BDT and DT classifiers on the training data. Figures 6-2 was plotted based on information obtained from Tables 6-4 to 6-6 and 6-10 to 6-12 illustrating the performance of the BDT and DT classifiers on the testing data. In these plots, the x-axis represents the percentage of fraud rate in the training set whereas the y-axis represents the FN rate of the classifiers.Boosted

Decision Tree Evaluation on Training Data

Table 6.1: Evaluation on Training Data

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.114	9951	98	1429	1927	0.99	0.575	0.425	0.01
33: 67	0.152	6455	248	1276	2080	0.964	0.619	0.38	0.036
50: 50	0.185	2938	425	821	2535	0.874	0.755	0.245	0.126

Table 6.2: Decision Tree Evaluation on Testing Data Part 2

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.127	9881	168	1538	1818	0.983	0.542	0.458	0.017
33: 67	0.167	6530	173	1503	1853	0.975	0.553	0.447	0.025
50: 50	0.197	3025	338	987	2369	0.9	0.706	0.294	0.1

Table 6.3: Decision Tree Evaluation on Testing Data Part 3

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.145	9941	108	1831	1525	0.97	0.454	0.543	0.03
33: 67	0.172	6430	273	1455	1901	0.96	0.567	0.433	0.04
50: 50	0.22	2932	431	1046	2310	0.872	0.689	0.311	0.128

Decision Tree Evaluation On Testing Data

Table 6.4: Boosted Decision Tree Evaluation On Testing Data Part 1

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.137	4671	385	503	906	0.924	0.643	0.357	0.076
33: 67	0.16	4481	575	952	952	0.886	0.676	0.324	0.114
50: 50	0.232	3806	1250	248	1161	0.753	0.824	0.176	0.247

Table 6.5: Boosted Decision Tree Evaluation On Testing Data Part 2

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.129	4696	360	471	938	0.928	0.666	0.334	0.0712
33: 67	0.133	4640	416	442	967	0.918	0.686	0.314	0.082
50: 50	0.245	3750	1306	281	1128	0.742	0.80	0.20	0.285

Table 6.6: Boosted Decision Tree Evaluation On Testing Data Part 3

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.12	4800	256	571	892	0.949	0.633	0.367	0.051
33: 67	0.138	4582	474	419	990	0.907	0.702	0.298	0.093
50: 50	0.259	3638	1418	297	1112	0.72	0.79	0.21	0.28

Decision Tree Evaluation on Training Data

Table 6.7: Evaluation on Training Data (All features are considered)

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.135	9679	370	1446	1910	0.964	0.569	0.431	0.036
33: 67	0.173	6402	301	1438	1918	0.965	0.572	0.428	0.044
50: 50	0.206	2972	391	996	2360	0.884	0.703	0.297	0.116

Table 6.8: Evaluation on Training Data (card type is disregarded)

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.149	9903	146	1845	1511	0.958	0.451	0.549	0.042
33: 67	0.185	6150	553	1312	2044	0.918	0.61	0.39	0.082
50: 50	0.224	2923	440	1067	2289	0.87	0.682	0.318	0.130

Table 6.9: Evaluation on Training Data (POS &card are disregarded)

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.153	9874	175	1870	1486	0.96	0.443	0.557	0.04
33: 67	0.192	6181	522	1411	1945	0.922	0.58	0.42	0.078
50: 50	0.235	2836	527	1053	2303	0.844	0.686	0.314	0.156

Decision Tree Evaluation On Testing Data

Table 6.10: Evaluation On Testing Data (All features are considered)

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.144	4511	545	384	1025	0.893	0.728	0.272	0.107
33: 67	0.144	4541	515	418	991	0.899	0.704	0.296	0.101
50: 50	0.238	3798	1258	279	1130	0.752	0.802	0.198	0.248

Table 6.11: Evaluation On Testing Data (card type is disregarded)

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.119	4828	228	544	865	0.955	0.614	0.386	0.045
33: 67	0.165	4384	672	397	1012	0.867	0.719	0.281	0.133
50: 50	0.25	3729	1327	292	1117	0.738	0.793	0.207	0.262

Table 6.12: Evaluation On Testing Data (POS &card are disregarded)

Class distribution	Error rate	TN	FP	FN	TP	TN rate	TP rate	FN rate	FP rate
25: 75	0.118	4816	240	522	887	0.953	0.63	0.37	0.047
33: 67	0.155	4430	626	378	1031	0.876	0.732	0.268	0.124
50: 50	0.275	3572	1484	293	1116	0.71	0.8	0.2	0.29

To choose the most effective classifier using the appropriate class distribution, the performance of the classifiers To choose the most effective classifier using the appropriate class distribution, the performance of the classifiers should be analyzed. Figures 6-1 and 6-2 are used as the basis for the analysis. These Figures and the associated Tables demonstrate that the boosted decision trees (BDT) trained on 25:75 fraud/non-fraud distribution attained TN rates of 99% and 92.4% and FN rates of 42.5% and 35.7% on the training and testing data, respectively. The comparative decision tree (DT) classifier attained TN rates of 96.4% and 89.3% and FN rates of 43.1% and 27.2%.These Figures also indicate that FN rate decreases as the minority cases increase in the training data and is lowest at 50:50 class

distribution. As these Figures show, BDT classifier trained on 50:50 class distribution attained the TN rates of 87.4% and 75.3% and FN rates of 24.5% and 17.6% on training and testing data, respectively. This classifier attained the lowest FN rate (i.e., 24.5% and 17.6%) among all the other classifiers.The desired classifier is the one that can identify as many legitimate transactions (TNs) as possible while not misclassifying the fraudulent transactions (FNs), otherwise significant losses will occur. Based on this goal, BDT classifier trained on 50:50 distribution by having 17.6% FN rate on testing set, appears to be the most predictive classifier among all the other classifiers constructed in this study.

6.3 PREDICTION OF NEW CASES

Once the most effective classifier was found, its quality could be further assessed by examining its prediction accuracy on the new cases. New cases are the ones that have not been used in the training or testing procedure and were referred to as the case set. A set of 250 transactions from legitimate accounts and another set of 222 transactions from fraudulent accounts were used for this evaluation.For prediction on new cases, BDT classifier was used. When using the classifier, an interactive window asks for the values of the data attributes associated with the example. All the values will be entered manually. The features requested, and the order in which they are requested, depend on the classifier itself. For instance, the classifier may ask for the value of ‘dollar amount’ or ‘merchant country’ as the first attribute and then it will ask for the second attribute which can be ‘card type’, or any other feature. After all the necessary attribute values have been entered, the most probable class is shown with a probability value. This value is a number, in the range of 0 to 1, associated with the prediction of Fraud (Y) / Non-fraud (N) class. Two examples of this prediction are shown in Figures 6.3 and 6.4.To predict the class of each transaction, the required values were entered interactively and the predicted class along with the probability value for each prediction are shown in Tables D.l and D.2 of Appendix D. These Tables contain the transaction features, their values, predicted class, the probability value associated with this class prediction, and the correct class for each transaction.

6.4 CONCLUDING REMARKS

The performance of BDT classifiers, trained on 25:75 and 50:50 class distribution, respectively. As these results demonstrate, the classifier trained on 25:75 class distribution has the FN rate of 49.7% on the classification of fraudulent transactions (by missing half of the fraudulent cases) whereas the comparative classifier trained on 50:50 distribution has the FN rate of 26.8%. This comparison reaffirms that BDT classifier trained on 50:50 distribution is the higher performance classifier for the prediction of new cases.One question that might arise is what happens if the class distributions of 60:40 (60% fraud cases in training set) or 75:25 (75% fraud cases in training set) is considered for the datasets? As stated before, the fraud dataset was rather small, therefore, it was not possible to form these distributions and explore their effects on the performance of the system. Yet other researchers [6] have examined these distributions for their fraud detection analysis and their results show that the existence of fraud cases higher than 50% in the training set degraded the performance of the classifier and based on their results, they concluded that 50:50 class distribution is the suggested distribution for the construction of higher performance classifiers.

7. CONCLUSION

Fifty Four experiments were conducted to determine the most predictive classifier. These experiments were performed based on several variations and combinations of features and training class distributions. The evaluation of classifiers, constructed from different sets of experiments, was different on training and testing data confirming that significant attention has to be paid in the class distribution design of the training sets. The performance metrics considered for this analysis were True Negative (TN) and False Negative (FN) rates. The BDT classifier trained on 50:50 class distribution attained a TN rate of 87.4% and 75.3% and FN rate of 24.5% and 17.6% on training and testing data, respectively. Based on this performance, this classifier considered being the most predictive classifier for this study by having the lowest possible FN rate among all the other classifier constructed in this analysis. This analysis reaffirms the importance of training class distribution in the design of the effectiv classifiers. This study shows that increasing the number of minority instances in the training data will produce classifiers with improved performance. It also shows that increasing the number of majority instances in the training data will produce classifiers that are adept at classifying the majority of transactions as legitimate and as a result, these classifiers classify a large number of fraudulent cases as legitimate leading to very high FN rate. The performance of the BDT classifiers on the prediction of new cases was also examined. This analysis showed that the classifier trained on 25:75 distribution of fraud/legitimate transactions attained the TN rate of 98.8% in the prediction of legitimate cases. However, the performance of this classifier degraded on the identification of fraudulent cases so that the classifier identified half of the fraudulent transactions as legitimate, attaining a FN rate of 49.7%. The degradation in performance makes the system unusable because missed fraud cases are very costly. The classifier trained on 50:50 distribution had lower TN rate (92% against 98.8%) on the prediction of legitimate transactions, however, its FN rate on the prediction of fraud cases was very much lower (26.8% against 49.7%) than the comparative classifier. This analysis reaffirms that classifier trained on 50:50 class distribution is more predictive for the evaluation of new cases. The other important factor which may have a serious impact on the performance of the classifiers was the limitations of the data sets. The most important limitations were rather small fraud database and the lack of FDS scores associated with the flagged transactions.

These scores are an indication of some patterns of behavior in the datasets and contain valuable information. The result of the experiments conducted on the variations of features revealed that the classifiers trained on all features performed much better than the ones train while disrega rding some features such as POS and card type. Based on these empirical results one would expect that if the FDS scores were provided, they would contribute important information thus leading to better performance. There was no information available on the cost of investigation a ssociated with every case created by the FDS, therefore, the savings from the use of the system trained could not be estimated. Based on the observed results on the prediction of new cases, one could expect that this approach may reduce the volume of personal investigations leading to pot entially significant savings for the FI. Currently, due to the high volume of false positives flagge d by the FDS, the FI has set a rather high threshold for this system. Therefore, there are cases that are fraudulent but are being missed (FN) by the FDS. By instituting a postprocessor system suc h as the ML, the FI has the option of lowering the threshold and allowing FDS to flag more cases for investigation. Another important point is the prevention of unnecessary disturbance of the cus tomers which may lead to customer dissatisfaction. In summary, pattern recognition for legitima te/fraud occurrences is inherently complex and since legitimate cardholders'/ fraudsters' pattern s of behavior evolve over time, this study is a basis for further research. Overall this study demo nstrates that the approach employed in this research, has a very good potential of identifying the legitimate transactions from the fraudulent ones. In Future Work, The potential of the trained system for the identification of legitimate transactions from the fraudulent ones, flagged by FDS, is promising but there is a need for the enhancement of its predictive accuracy. Some of the most important recommendations for future research, that can explore the possibilities of enhancements on the prototype and its potential deployment in credit card fraud detection are listed below:

- Learning systems do the best they can with what they are given. It is quite possible that revising or adding new features may lead to much better performance for the same learning method [WEIS91 j. Sufficient and representative data are the foundation of all learning systems and this study showed that using all features produced classifiers with better performance. Therefore, for further improvement on the trained system, data requirements must be fulfilled. In this respect, two major requirements for further analysis are FDS scores and a larger fraud dataset.

- To form datasets with higher minority instances (i.e., 60:40, 70:30, etc. of fraud/non fraud cases) in order to explore the effect of class distribution on the performance of the classifiers and based on this evaluation to choose the most predictive classifier for use.
- There are different learning techniques that can be applied to the same sample data. For a given application, some learning systems may be better than others. In general, there is no guarantee that any of these methods work or that any single method is necessarily the best. In this study one prominent learning software (See5) was utilized. The two other well-known software, namely, CART [2] and RIPPER [10] should also be examined. These software have been applied in real world problems such as credit card fraud detection and they have shown impressive results. By employing different techniques the performance can be measured and the algorithm which yields the best performance can be selected.
- As discussed before, fraud environment is dynamic, therefore, the system being designed must be adaptive to changing fraud environment.

REFERENCES

[1] S. Y. S. Navanshu Khare, "Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 1–4, pp. 23–27, 2018.

[2] S. Bhatia, R. Bajaj, and S. Hazari, "Analysis of Credit Card Fraud Detection Techniques," *Int. J. Sci. Res.*, vol. 5, no. 3, pp. 1302–1307, 2016.

[3] C. P. Lim, M. Seera, A. K. Nandi, K. Randhawa, and C. K. Loo, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, no. 11, pp. 14277–14284, 2018.

[4] M. Zareapoor and P. Shamsolmoali, "Application of credit card fraud detection: Based on bagging ensemble classifier," *Procedia Comput. Sci.*, vol. 48, no. C, pp. 679–685, 2015.

[5] Y. Sahin *et al.*, "Adaptive Machine Learning for Credit Card Fraud Detection Declaration of Authorship," *Decis. Support Syst.*, vol. 50, no. 3, pp. 103–106, 2011.

[6] I. Trivedi, M. M, and M. Mridushi, "Credit Card Fraud Detection," *Ijarcce*, vol. 5, no. 1, pp. 39–42, 2016.

[7] M. R, "Fraud Detection using Supervised Learning Algorithms," *Ijarcce*, vol. 6, no. 6, pp. 6–10, 2017.

[8] S. patel and S. Gond, "Supervised Machine (SVM) Learning for Credit Card Fraud Detection," *Int. J. Eng. Trends Technol.*, vol. 8, no. 3, pp. 137–139, 2014.

[9] D. Choi and K. Lee, "Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System," *IT Converg. Pract.*, vol. 5, no. 4, pp. 12–24, 2017.

[10] I. Trivedi, M. M, and M. Mridushi, “Credit Card Fraud Detection,” *Ijarcce*, vol. 5, no. 1, pp. 39–42, 2016.

[11] C. Whitrow, D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams, “Transaction aggregation as a strategy for credit card fraud detection,” *Data Min. Knowl. Discov.*, vol. 18, no. 1, pp. 30–55, 2009.

[12] L. S. Raghavendra Patidar, “Credit Card Fraud Detection Using Neural Network,” *India Int. J. Soft Comput. Eng.*, vol. 9194145521, no. 1, pp. 13–14, 2011.

[13] P. K. Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo, “Distributed Data Mining in Credit Card Fraud Detection,” *IEEE Intell. Syst. Their Appl.*, vol. 14, no. 6, pp. 67–74, 1999.

[14] A. Mane, V. Dighe, R. Gawali, S. Sabale, and S. Gudadhe, “Location based Service and Health Monitoring System for Heart Patient using IoT,” *Mon. Peer Rev. Journal) Website www.ijircce.com*, vol. 5, no. 11, pp. 1785–1793, 2017.

[15] H. Lee *et al.*, “Feature selection practice for unsupervised learning of credit card fraud detection,” *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 2, pp. 408–417, 2018.

[16] P. R. Shimpi, “Survey on Credit Card Fraud Detection Techniques,” *Int. J. Eng. Comput. Sci.*, 2016.

[17] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, “Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning,” *Inf. Fusion*, vol. 10, no. 4, pp. 354–363, 2009.

[18] P. K. Chan and S. J. Stolfo, “Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection,” *Proc. Fourth Int. Conf. Knowl. Discov. Data Min.*, pp. 164–168, 1998.

[19] S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick, “Machine Learning Techniques for Fraud Detection,” *Proc. 1st Int. naiso Congr. neuro fuzzy Technol.*, no. January, pp. 261–270, 2002.

[20] S. Aihua, T. Rencheng, and D. Yaochen, “Application of classification models on credit card fraud detection,” *Proc. - ICSSSM’07 2007 Int. Conf. Serv. Syst. Serv. Manag.*, no. March, 2007.

[21] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, “Association rules applied to credit card fraud detection,” *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 3630–3640, 2009.

[22] L. Xu, R. Egawa, H. Takizawa, and H. Kobayashi, “A network clustering algorithm for Sybil-Attack resisting,” *IEICE Trans. Inf. Syst.*, vol. E94–D, no. 12, pp. 2345–2352, 2011.

[23] Ghosh and Reilly, “Credit card fraud detection with a neural-network,” *Proc. Twenty-Seventh Hawaii Int. Conf. Syst. Sci. HICSS-94*, pp. 621–630, 1994.

[24] R. J. Bolton and D. J. Hand, “Statistical Fraud Detection: A Review,” *Source Stat. Sci. Stat. Sci.*, vol. 17, no. 3, pp. 235–249, 2002.

[25] V. Zaslavsky and A. Strizhak, “Credit Card Fraud Detection Using Self-Organizing Maps,” *Inf. Secur. An Int. J.*, vol. 18, pp. 48–63, 2013.

[26] S. Y. S. Navanshu Khare, “Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models,” *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 1–4, pp. 23–27, 2018.

[27] L. Delamaire, H. Abdou, and J. Pointon, “Credit card fraud and detection techniques: a review,” *Banks Bank Syst.*, vol. 4, no. 2, pp. 57–68, 2009.

[28] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, “Data mining for credit card fraud: A comparative study,” *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011.

[29] A. Dal Pozzolo, O. Caelen, Y. A. Le Borgne, S. Waterschoot, and G. Bontempi, “Learned lessons in credit card fraud detection from a practitioner perspective,” *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4915–4928, 2014.

[30] A. Pumsirirat and L. Yan, “Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 1, pp. 18–25, 2018.

APPENDIX A

CARDHOLDERS GLOSSARY

A.1 NOMENCLATURE

ABM - Automated Banking Machine

Attributes - See features

Authorization - To be able to make a purchase by credit cards, cardholder's FI must authorize the transaction from the central computer.

Authorization Log - FI system keeps a record of all the authorizations that pass through its mainframe in a database called "authorization log'' for future reference.

Bias - A preference for one hypothesis over another. Since in most learning situations there are a variety of possible consistent hypotheses, all learning algorithms have some sort of bias.

Cardholder File - All the information related to the cardholder is kept in this file for accounting purposes.

Card Identification Device (CID) - Special security feature included in the magnetic stripe of American Express to counteract the counterfeiting process.

Card Issuers (CIs) - Institutions that issue credit cards.

Card Verification Value (CW) - Special security feature included in the magnetic stripe of VISA to counteract the counterfeiting process.

Card Verification Code (CVC) - Special security feature included in the magnetic stripe of MasterCard to counteract the counterfeiting process.

CBA - Canadian Bankers Association.

Classification - To assign a specific class to a case.

Classifier - A decision-making system that classifies the class of cases based on the pattern instances it has learned, is called a classifier. The simplest way of representing a classifier is as a black box, which produces a decision for every admissible pattern of data that is presented to it. It accepts a pattern of data as input, and produces a decision as output.

Concept - A classification rule that partitions a domain into two parts: those instances that satisfy it and those that do not satisfy it.

Concept Learning - Inferring a Boolean-valued function from training examples of its input.

Confusion Matrix - A matrix which pinpoints the kinds of errors made in the analysis. This matrix shows the detail breakdown of correctly and incorrectly classified cases.

Credit Limit - The restricted maximum amount assigned by the FIs on each card issued to a cardholder. Any credit in excess of such limit will require the issuer's authorization to enable any transaction above that limit.

Decision Trees -

A simple structure for inductive learning. Given an instance of the problem, specified by a set of featu res and theirvalues, a decision tree returns a "yes" or"no" decision aboutthe instanceTherefore, decisi on trees are Boolean classifiers. Each branching node in the tree represents a test on some aspect of th e instance.

Delinquent - In cases where the cardholder payment is less than the minimum amount, the credit rating of the cardholder is affected and the cardholder is considered delinquent.

Error Rate - The most common measure for evaluating the performance of classifiers is error rate (1- accuracy. This ratio measures the percentage of incorrectly classified instances and has the implicit assumption that each error is equally important.

False Negative - When the system misses a fraudulent transaction.

False Positive - When the system flags a legitimate transaction as fraudulent.

Features - The sets of potential observations relevant to a particular problem are referred to as features. Features are also known by other names such as 'attributes', and 'variables'.

Financial Institutions (FIs) - Banks, credit unions, trust companies, major retailers, etc.

Floor Limit - There are merchants who have assigned a floor limit and purchases below that amount could be authorized by the merchant and need not be authorized through cardholder's FI system. The limit set depends on the kind of business, the store location, type of merchandise or service and other factors. Any value in excess of the floor limit requires the authorization of the card issuer.

Fraud Analyst - Human experts employed and trained by the FIs to follow up and investigate suspicious transactions in order to detect fraud.

Inductive Learning -

Inductive learning is a kind of learning in which, given a set of instances the system tries to estimate or create an evaluation function. Most inductive teamingis supervised learning, in which ex

amples are provided with classification. More formally, an example is a pair of (x, ff x) where x is the input and f(x) is the output of the function applied to x. The task of induction is, given a set of example s of f find a hypothesis h that approximates f.

.

Learning - An approach to improve problem solving through experience. It is "an increase in knowledge when knowledge is knowledge in principle."

Machine Learning - Class of programs and algorithms that improve through experience. These programs search over a large space of hypothesis to find the one that best fits to the characteristics of the training data.

Magnetic Stripe - A dark, machine-readable stripe on the back of the plastic cards for storing card holder information.

Mainframe - Central computer of FIs in charge of a number of important activities such as processing the incoming transactions for authorization, record tracking, issuing monthly statements, and so on.

Merchant File - For accounting purposes FIs keep merchant records and information in a file called "merchant file".

Neural Networks - A class of knowledge-based models in AI.

Negative File - Due to the fact that having a copy of each VISA cardholder in the FI's system is not practical, all the card numbers that have been considered fraudulent internationally, are included in a file called 'negative file' which is updated quite frequently with the occurrence of new fraud cases.

Noise - When there is contradictory information in the data such as two or more examples with the same descriptions (in terms of the attributes) with different classifications. In other words, examples might have exactly the same description but a different classification is assigned to them. This means that some of the data are incorrect. If this happens then the decision tree learning algorithm must fail to find a decision tree consistent with all the examples. This happens when data is labeled incorrectly (e.g., the examples were positive but were labeled as negative)

Off-line - When the system is not connected to a computer or data communications network.

On-line Authorization - When authorization of a transaction uses equipment which is connected to a computer or data communications network and is carried out in real time.

Personal Identification Number (PIN) - The security code assigned to the card to be used in an Automated Banking Machine (ABM).

Point of Sale (POS) - Location at a merchant where a customer makes a purchase.

Point of Sale (POS) Terminal - A machine placed in a merchant location which is connected to the FI's on-line authorization system via a modem, designed to authorize, record and forward data for each transaction.

Posted Transaction File - This file keeps a record of all the current transactions that a cardholder has made and as yet has not posted to the statement This file calculates the current balance of every account, keeps a record of them and at the end of the month this information will be posted to the cardholder's statement.

Smart Cards - Smart cards feature a microprocessor memory chip as well as data encoded on its magnetic strip.

Stand In Processing - In occasions when FI's mainframe is non-functional for authorization, the Tandem does the authorization considering two criteria: (1) the 'negative file', (2) an assigned floor limit.

Statement - A list of all the cardholder's transactions during one accounting period.

Statement File - This file is used for keeping track of the statement balances and payments. At the end of the cardholder's cycle the accumulated transactions will be sent to this file. The monthly statement for the cardholder is printed out of this file.

Supervised Learning - Any situation in which both the inputs and outputs of a component can be observed.

Swipe Machine - Same as the POS machine (see Point of Sale).

Tandem - A non-stop, central computer being used by FIs to process all the incoming transactions and route them to proper place for authorization. Also in the absence of mainframe it does the 'stand in' processing.

Test Set - A set of instances and their classifications used to test the accuracy of a learned system. The training set is used to create the classifier. The test set is used to validate the performance of the classifier.

Training Set - A training set is a set of problem instances (described as a set of features and their values), together with a classification of the instance. Training sets are used in supervised learning.

Transaction - A cardholder makes a purchase using a credit card.

True Negative - When the transaction is legitimate and normal. True Positive - When the transaction is fraudulent and system hits it.

Unsupervised Learning - When there is no information about what the correct outputs are. Unsupervised learners can learn to predict future percepts based on present ones, but cannot leam which actions to take without a utility function.

Voice Authorization - There are merchants who do not have POS machines and have to call their FI and ask for authorization.

A - Transaction authorized

D - Transaction declined

K - Card keyed

S - Card swiped

R - Transaction referred to FI staff

P - Card has to be picked up

N - Transaction is legitimate

Y - Transaction is fraud

FP - False Positive

FN - False Negative

TP - True Positive

TN - True Negative

ML - Machine Learning

NN - Neural Network

CV - Cross Validation

DT - Decision tree

BDT- Boosted decision tree.