IMPROVING ACCURACY IN CREDIT CARD FRAUD DETECTION USING EFFICIENT DATA CLASSIFICATION

A Dissertation Proposal In partial fulfilment of the Requirement for the Award of the of Degree

MASTER OF TECHNOLOGY
In
COMPUTER SCIENCE & ENGINEERING

SUBMITTED BY

TAMANNA CHOUHAN

(11604887)

UNDER THE GUIDANCE OF

MR. RAVI KANT SAHU

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab (India)

December 2017

INDIA'S LARGEST UNIVERSITY
LOVELY PROFESSIONAL UNIVERSITY
Transforming Education, Transforming India

TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering

Program: P172::M.Tech. (Computer Science and Engineering) [Full Time]

COURSE CODE : CSE548 **REGULAR/BACKLOG :** Regular **GROUP NUMBER :** CSERGD0046

Supervisor Name : Ravi Kant Sahu **UID :** 16920 **Designation :** Assistant Professor

Qualification : ________________________ **Research Experience :** ________________________

SR.NO.	NAME OF STUDENT	REGISTRATION NO	BATCH	SECTION	CONTACT NUMBER
1	Tamanna Chouhan	11604887	2016	K1637	9805891878

SPECIALIZATION AREA : Networking and Security **Supervisor Signature:** ________________

PROPOSED TOPIC : Improving Accuracy in Credit Card Fraud Detection Using Efficient Data Classification

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	**Parameter**	**Rating (out of 10)**
1	Project Novelty: Potential of the project to create new knowledge	7.60
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	7.80
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	7.80
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	7.80
5	Social Applicability: Project work intends to solve a practical problem.	8.00
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	7.40

PAC Committee Members		
PAC Member 1 Name: Prateek Agrawal	UID: 13714	Recommended (Y/N): Yes
PAC Member 2 Name: Deepak Prashar	UID: 13897	Recommended (Y/N): Yes
PAC Member 3 Name: Raj Karan Singh	UID: 14307	Recommended (Y/N): NA
PAC Member 4 Name: Pushpendra Kumar Pateriya	UID: 14623	Recommended (Y/N): Yes
PAC Member 5 Name: Sawal Tandon	UID: 14770	Recommended (Y/N): NA
PAC Member 6 Name: Aditya Khamparia	UID: 17862	Recommended (Y/N): Yes
PAC Member 7 Name: Anupinder Singh	UID: 19385	Recommended (Y/N): NA
DAA Nominee Name: Kuldeep Kumar Kushwaha	UID: 17118	Recommended (Y/N): Yes

Final Topic Approved by PAC: **Improving Accuracy in Credit Card Fraud Detection Using Efficient Data Classification**

Overall Remarks: Approved

PAC CHAIRPERSON Name: 11024::Amandeep Nagpal **Approval Date:** 04 Nov 2017

ABSTRACT

The number of customers for the credit cards (CC) has grown of the last decade. These cards have aggravated the cashless systems of payments as well as alleviated the use of cash credit, which is termed as a short-term continuous loan. The credit cards are known to increase the purchasing power of citizens, and let them meet their daily needs, gadgets, etc. The number of CC (credit card) frauds has increased with the increase in number of credit cards. The unethical use of credit cards by hackers or credit cards users unwilling to pay back the amount are known as the major credit card frauds.

The credit card frauds can be detected by evaluating the CC purchasing patterns using the historical data in order to detect the frauds. This data evaluation can help the banks or other organizations offering credit cards to minimize their losses due to the credit card frauds. The historical data evaluation with the current purchasing patterns requires the statistical modelling, which can automatically evaluate the fraudulent patterns and alarm the banks about the transactions. This helps the banks for early detection of the frauds, where they can easily eliminate the CC frauds by declining the suspected transactions.

DECLARATION

I hereby declare that the research work reported in the dissertation proposal entitled **"IMPROVING ACCURACY IN CREDIT CARD FRAUD DETECTION USING EFFICIENT DATA CLASSIFICATION"** in partial fulfilment of the requirement for the award of Degree for Master of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor Mr. Ravi Kant Sahu. I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University's Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

Tamanna Chouhan

(11604887)

SUPERVISOR’S CERTIFICATE

This is to certify that the work reported in the M. Tech Dissertation proposal entitled **“IMPROVING ACCURACY IN CREDIT CARD FRAUD DETECTION USING EFFICIENT DATA CLASSIFICATION”**, submitted by **Tamanna Chouhan** at **Lovely Professional University, Phagwara, India** is a bonafide record of her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Ravi Kant Sahu

Date:

1) **Concerned HOD:**

   HoD’s Signature: ________________

   HoD Name: ____________________

   Date: ___________________

2) **Neutral Examiners:**

   External Examiner

   Signature: _______________

   Name: __________________

   Affiliation: ______________

   Date: ___________________

   Internal Examiner

   Signature: _______________

   Name: __________________

   Date: ___________________

ACKNOLOGEMENT

I would like to convey my most heartfelt and sincere gratitude to my mentor Mr. Ravi Kant Sahu of Lovely Professional University, for his valuable guidance and advice. His willingness to motivate me contributed tremendously to achieve the goal successfully.

I would also like to thanks dear God and my parents who has always inspired and encouraged me to achieve my goal successfully.

Tamanna Chouhan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Misrepresentation alludes to getting products/administrations and cash by illicit way. Extortion manages occasions which include criminal intentions that, for the most part, are hard to recognize. Charge cards are a standout amongst the most mainstream goal of extortion yet not alone. Charge card extortion, a colossal term for burglary and misrepresentation submitted or any comparable instalment instrument as a fake asset of assets in an exchange. Charge card misrepresentation has been growing issue in the Master card business. Recognizing Visa misrepresentation is a troublesome undertaking when utilizing typical process, so the advancement of the charge card extortion identification models has happened to significance whether in the scholastic or business associations right now. Besides, part of extortion has been changed abruptly amid the most recent couple of decades alongside headway of technologies.

1.1 Credit Card Fraud

Credit Card Misrepresentation is one of the greatest dangers to business and business foundations today. Just, Master card Misrepresentation is characterized as, "when an individual uses another individuals“ Visa for individual utilize while the proprietor of the card and also the card backer don't know about the thing that the card is being utilized." various frameworks/models, process and preventive measures will stop Master card extortion and lessen monetary dangers. Banks and Visa organizations have accumulated a lot of Master card account exchanges.

The Charge card is a plastic card issued to number of clients as one of the method of instalment. It enables cardholders to buying merchandise and enterprises in light of the cardholder's guarantee.

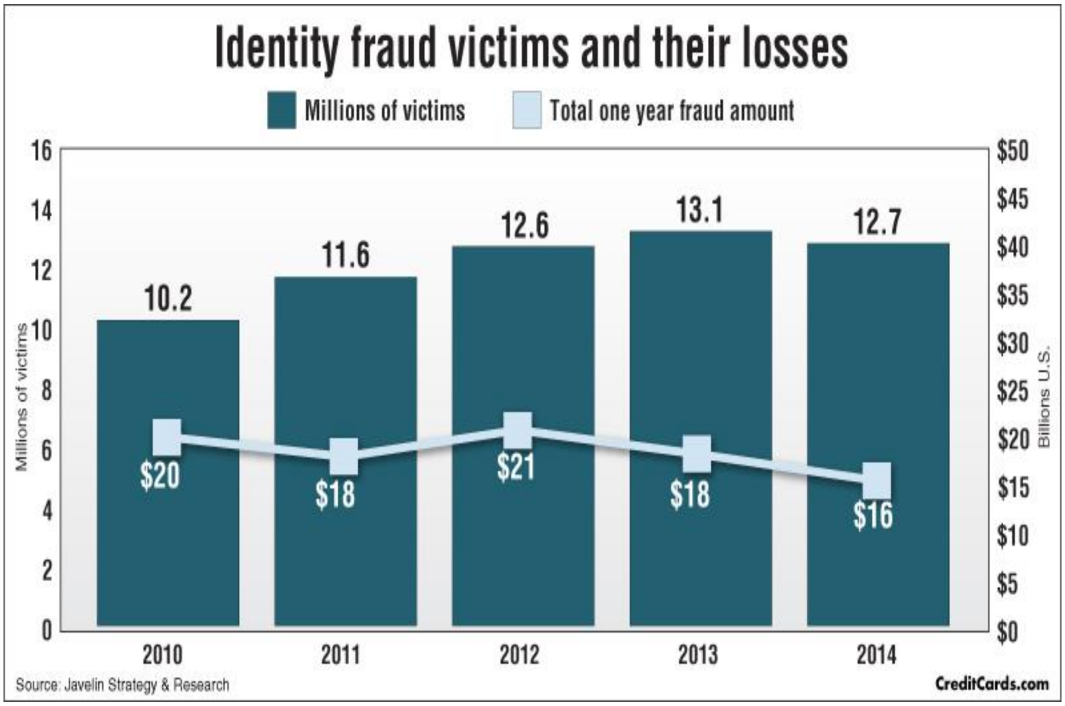


Figure 1.1: Identity Fraud Victims and their losses

In China, Visa clients are developing quickly, yet just a not very many Master card holders utilize charge cards for paying for everyday buy similarly with certainty and a suspicion that all is well and good. Reason is, Visa holder has no enough certainty to trust upon the instalment framework. Secure credit administrations of banks and improvement of E-business a solid extortion identification framework is basic to help safe Visa use, Misrepresentation location in light of breaking down existing buy information of cardholder (current spending conduct) is a promising path for decreasing the rate of Visa cheats.

As there is constrained measure of information with the exchanges being trusted, for instance, exchange sum, shipper classification code (MCC), acquirer number and date and time, address of the trader. Different procedures in Learning Disclosure, for example, choice tree, neural system and case based thinking have extensively been utilized for framing a few misrepresentation location frameworks/models. These methods more often than not require sufficient number of typical exchanges and misrepresentation exchanges for learning extortion designs. Be that as it may, the proportion of false exchanges to its typical exchanges is low greatly, for an individual bank.

1.2 Outlier Detection Methods

Outlier detection is the process to discover the patterns out of the data. In simplified words, the outliers are the data patterns not following the normalcy rules, and stands out of the dataset. The outlier detection is used to discover the special patterns to understand the upper or lower bound variations in the data distribution. The normal distribution can be described with the following graph, which elaborates the data points around the mean and standard deviation.

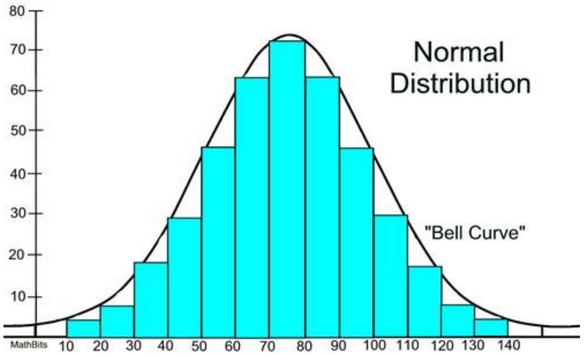


Figure 1.2: Describing the normal distribution

The normal distribution is the generalized distribution of data, which theoretical meant to be present in most of the datasets. In real-time, the datasets are slightly deviated towards one side, which is known as skewness. Generally, nearly all of the datasets are found distributed in the near normalized manner, which is generally considered as the normal distribution.

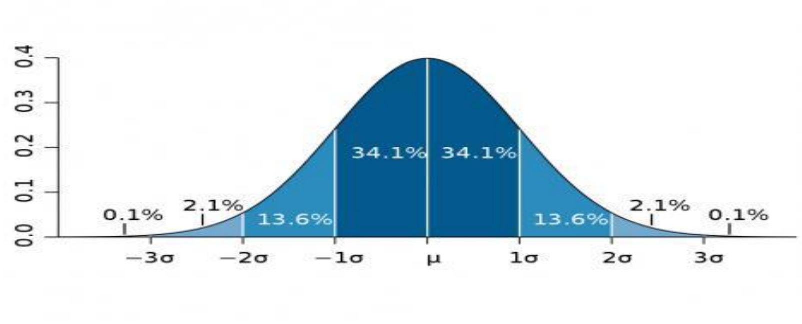


Figure 1.3: The normal distribution popular division around the standard deviation. Theoretically, the data divisions under the normal distribution is divided according to the presented percentages in the above figure

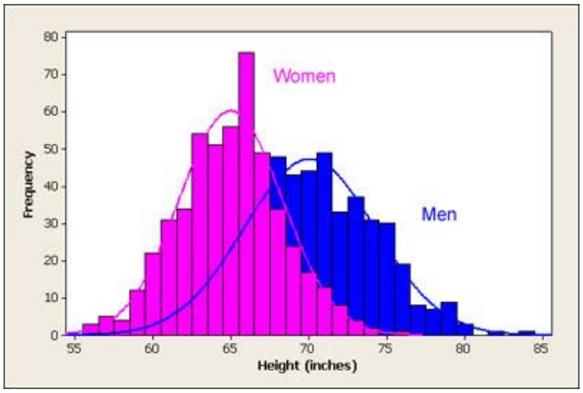


Figure 1.4: Example of outlier detection

The outlier detection process can be described from the figure 3, where the number of women with height of 66 inches in exceptionally higher. This phenomenon is known as the outlier or anomaly, and used to explain the interesting facts about data. In the case of credit cards, the outlier detection is used to discover the fraudulent patterns, and used to minimize the economic losses caused by the CC frauds.

1.2 Different Types of Credit Card Fraud

Misrepresentation discovery frameworks come into situation when the fraudsters surpass the extortion aversion frameworks and begin false exchanges. Alongside the advancements in the Data Innovation and upgrades in the correspondence channels, misrepresentation is spreading everywhere throughout the world with aftereffects of vast measure of false misfortune.

Anderson (2007) has recognized and depicted the distinctive sorts of extortion. Charge card cheats can be continue in a wide range of courses, for example, basic burglary, fake cards, Never Got Issue (NRI), application misrepresentation and on the web/Electronic extortion (where the card holder is absent). Master card misrepresentation recognition is appallingly troublesome, yet in addition regular issue for arrangement.

1.3.1 Card not present transaction (CNP)

In the event that a card isn't physically present when a client makes a buy, the shipper must depend on the cardholder, or somebody indicating to be thus, introducing card data in a roundabout way, regardless of whether via mail, phone or over the Internet.

1.3.2 Identify Theft

The Identity Theft is divided into two classifications:

- **Application fraud**

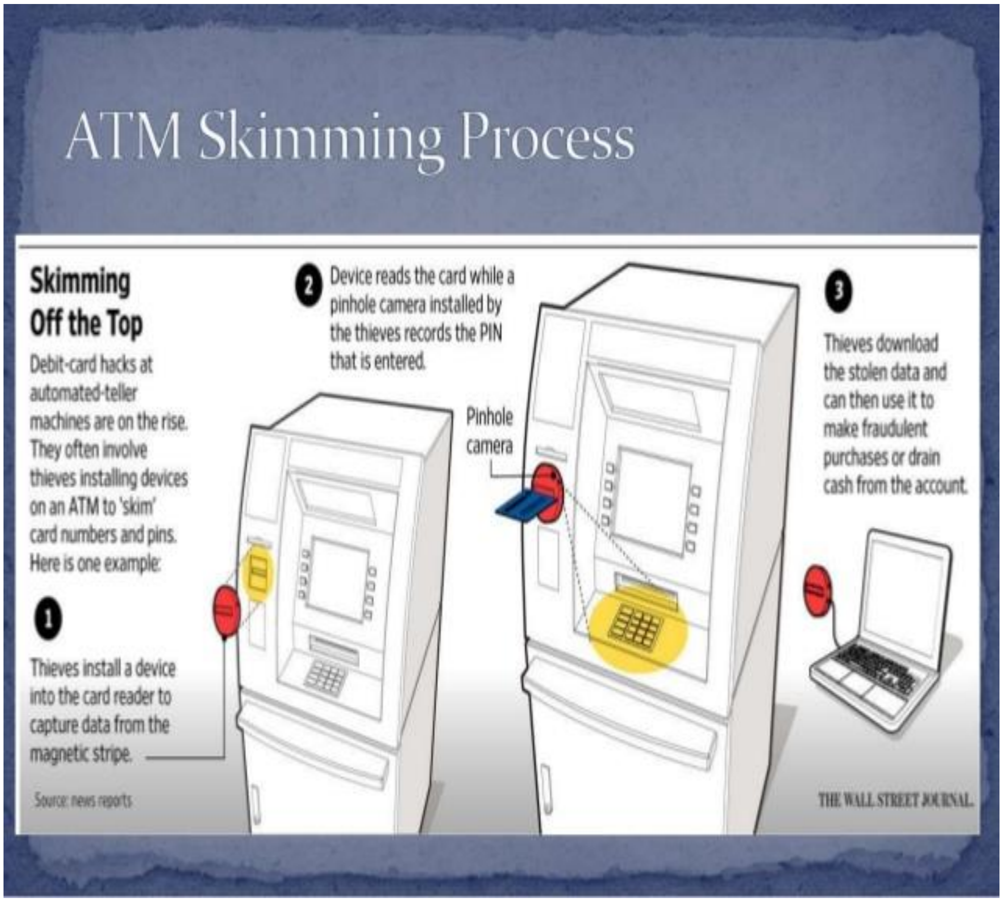
Application fraud happens when a man utilizes stolen or counterfeit archives to open a record in someone else's name. Culprits may take records, for example, service bills and bank explanations to develop helpful individual data

- **Account takeover**

This theft happens when an illegal poses as an intelligent customer, gains control of an account and then makes unofficial transactions

1.3.3 Skimming

An electronic strategy for catching a casualty's close to home data utilized by personality criminals. The skimmer is a little gadget that outputs a Visa and stores the data contained in the attractive strip. Skimming can occur amid a true blue exchange at a business.



ATM Skimming Process

Skimming Off the Top

Debit-card hacks at automated-teller machines are on the rise. They often involve thieves installing devices on an ATM to 'skim' card numbers and pins. Here is one example:

1 Thieves install a device into the card reader to capture data from the magnetic stripe.

2 Device reads the card while a pinhole camera installed by the thieves records the PIN that is entered.

3 Thieves download the stolen data and can then use it to make fraudulent purchases or drain cash from the account.

Source: news reports

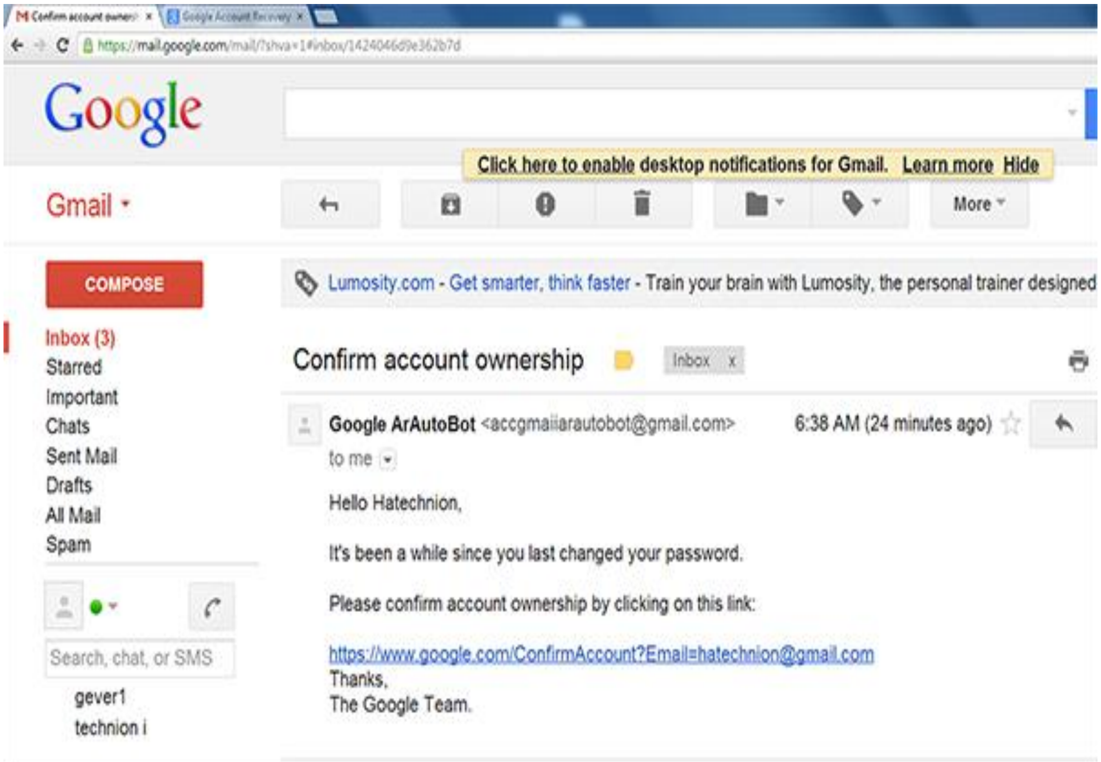
THE WALL STREET JOURNAL.

Figure 1.5: ATM skimming process

1.3.4 Phishing

Phishing email messages, sites, and telephone calls are intended to take cash. Cybercriminals can do this by introducing pernicious programming on your PC or taking individual data off of your PC. Phishing is a sort of social designing assault regularly used to take client information, including login certifications and Master card numbers. It happens when an assailant, taking on the appearance of a confided in element, hoodwinks a casualty into opening an email, text, or instant message. The beneficiary is then deceived into clicking a pernicious connection, which can prompt the establishment of malware, the solidifying of the framework as a component of a ransom ware assault or the noteworthy of delicate data.

An assault can have pulverizing comes about. For people, these incorporate unapproved buys, the taking of assets, or distinguish robbery.


Confirm account owners
Google Account Recovery
https://mail.google.com/mail/?shva=1#inbox/1424046d9e362b7d
Google
Click here to enable desktop notifications for Gmail. Learn more Hide
Gmail
More
COMPOSE
Lumosity.com - Get smarter, think faster - Train your brain with Lumosity, the personal trainer designed
Inbox (3)
Starred
Important
Chats
Sent Mail
Drafts
All Mail
Spam
Search, chat, or SMS
gever1
technion i
Inbox x

Confirm account ownership

Google ArAutoBot <accgmaiiarautobot@gmail.com> 6:38 AM (24 minutes ago)

to me

Hello Hatechnion,

It's been a while since you last changed your password.

Please confirm account ownership by clicking on this link:

https://www.google.com/ConfirmAccount?Email=hatechnion@gmail.com
Thanks,
The Google Team.

Figure 1.6: Real time example of Phishing attack

Credit card fraud is increasing rapidly, there are various techniques of detecting the credit card frauds some of them are support vector machine (SVM) AND Logistic regression. Based on these classification fraud transactions can we detected.

CHAPTER 2

REVIEW OF LITERATURE

Master card Fraud is one of the greatest dangers to business foundations today. Not with standing, to battle the misrepresentation viably, it is imperative to first comprehend the systems of executing a cheat. Master card fraudsters utilize countless usual methodology to submit extortion. In basic terms, Credit Card Fraud is characterized as "at the point when an individual uses another people's Visa for individual reasons while the proprietor of the card and the card backer don't know about the way that the card is being utilized". Further, the individual utilizing the card has no association with the cardholder or backer, and has no aim of either reaching the proprietor of the card or making reimbursements for the buys frantic. Charge card cheats are conferred in the following ways:

- A demonstration of criminal trickiness (misdirect with expectation) by utilization of unapproved account or potentially individual data.
- Illegal or unapproved utilization of record for individual pick up.
- Misrepresentation of record data to get products and additionally benefits.

Pallavi Kulkarni et.al [1] described that the Charge card is the all-around acknowledged way of reduction in money related field. With the rising number of clients over the globe, chances on use of MasterCard have additionally been expanded, where there is peril of taking MasterCard subtle elements and submitting cheats. Generally, machine learning territory has been creating calculations that have certain suppositions on fundamental dissemination of information.

- **Methodology Used**

In this paper author used the segment portrays a structure that tends to both the issues of stream information mining, i.e., idea float and class irregularity for Visa danger appraisal. For adjusting the information, the proposed system executes sub-gatherings with sacking alongside exchange blunder measures.

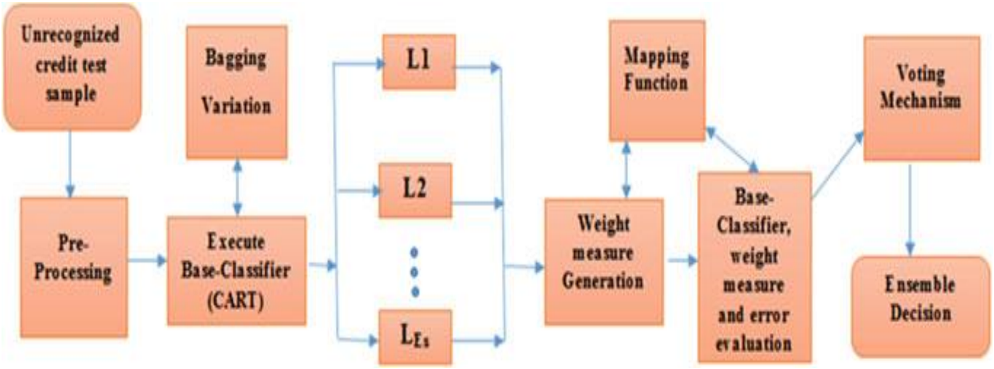

Unrecognized credit test sample
Pre-Processing
Bagging Variation
Execute Base-Classifier (CART)
L1
L2
L_{Es}
Mapping Function
Weight measure Generation
Base-Classifier, weight measure and error evaluation
Voting Mechanism
Ensemble Decision

Figure 2.1: Methodology used

Architecture of the proposed framework using logistic regression model Extent of future work is driving toward a brilliant strategy that will ideally dispose of downsides of the current research and furthermore evaluation of these approaches on substantial scale, true applications comprising of formal measurable investigations of these frameworks, on certain non-stationary situations like Gaussian dissemination floats.

Nuno Carneiro et.al [2] in this paper author explained about the credit card misrepresentation prompts billions of dollars in misfortunes for online dealers. With the advancement of machine learning calculations, scientists have been finding progressively advanced ways to distinguish extortion, however functional usage are once in a while revealed.

The author depict the improvement furthermore, arrangement of an extortion recognition framework in an extensive e-tail dealer. The paper investigates the blend of manual and programmed order, gives bits of knowledge into the total advancement process and thinks about various machine learning strategies.

This paper would thus be able to help scientist's .what's more, professionals to outline and actualize information digging based frameworks for misrepresentation location or comparative issues. This undertaking has contributed with a programmed framework, as well as with bits of knowledge to the misrepresentation examiners for enhancing their manual amendment process, which brought about a by and large predominant execution.

This work tended to the issue of extortion discovery for retailers working together on the web. The venture had the objective of outlining, creating and executing a hazard scoring framework (utilizing information mining procedures) at an e-tail dealer. This paper can help analysts and experts to outline what's more, execute information mining based frameworks, as it portrays the entire improvement process what's more, addresses down to earth execution issues.

Bertrand Lebichot et.al [3] discussed the worldwide card extortion misfortunes added up to 16.31 Billion US dollars in 2014. To recoup this colossal sum, mechanized Misrepresentation Location Frameworks (FDS) are utilized to deny an exchange before it is conceded. In this paper, we begin from a diagram based FDS named APATE, this calculation utilizes an aggregate deduction calculation to spread deceitful impact through a system by utilizing a constrained arrangement of affirmed false exchanges. APATE to fit to web based business field reality. This new strategy is evaluated on a three months genuine living web based business Visa exchanges informational index acquired from an expansive Master card backer.

In this paper, the author starts from an existing Fraud Detection Systems (FDS) APATE and bring several improvements. Another imagined additionally work is to present semi-directed learning on diagram investigation as well as in principle classified

Ashkan Zakaryazad et.al [4] said that the quick development in information catches and computational power has prompted an expanding centre around information driven research. Sofar, most of the examination is concentrated on prescient displaying utilizing measurable improvement, while benefit augmentation has been given less need. Ponder by adopting a benefit driven strategy to build up a benefit driven Fake Neural Network (ANN) .Characterization technique. In request to do this, author has first presented an ANN demonstrate with another punishment work which gives variable punishments to them is arrangement of examples considering their person significance (benefit of effectively arrangement and /or cost of mischaracterization) and after that author have considered expanding the aggregate net benefit. To produce individual penalties, author had altered the whole of squared mistakes (SSE) function by changing its esteems concerning benefit of each occurrence. We have executed distinctive renditions of ANN of which five of the female horse new ones contributed in this examination and two seat marks from important literature. In this investigation, an oval benefit based neural system has been proposed which makes the grouping considering all person expenses and benefits of each of the occurrences and subsequently boosts the aggregate net benefit caught from applying the grouping demonstrate.

Wen-Fang YU et.al [5] throws light on the expanding credit cards and developing exchange volume in china, credit card misrepresentation rises pointedly. Instructions to improve the identification and anticipation of visa misrepresentation turn into the concentration of hazard control of banks. This paper proposes a charge card misrepresentation identification show utilizing exception recognition in view of separation whole as per the rarity also, unpredictability of misrepresentation in charge card exchange information, applying exception mining into visa misrepresentation recognition. Trials demonstrate that this model is attainable and precise in identifying master card extortion.

This paper breaks down the plausibility of master card extortion identification in light of anomaly mining, applies exception discovery mining in light of separation total into master-card extortion identification and proposes this discovery methodology and its experimental process. Lastly this technique demonstrates exact in foreseeing false exchanges through exception mining copying investigation of charge card exchange informational collection of one certain business bank. The analysis demonstrates that anomaly mining can identify charge card extortion superior to irregularity identification in light of grouping when inconsistencies are far not as much as would be expected information. In the event that this calculation is connected into bank master card extortion recognition framework, the likelihood of extortion exchanges can be anticipated not long after visa exchanges by the banks. Furthermore, a progression of hostile to misrepresentation methodologies can be received to keep banks from incredible misfortunes earlier and lessen dangers.

K. R. Seeja et.al [6] described about the smart visa extortion discovery display for identifying misrepresentation from profoundly imbalanced and mysterious charge card exchange datasets. The class lop-sidedness issue is dealt with by finding legitimate and additionally extortion exchange designs for every client by utilizing continuous item set mining. A coordinating calculation is additionally proposed to discover to which design (legitimate or extortion) the approaching exchange of a specific client is nearer and a choice is made likewise. With a specific end goal to deal with the unknown nature of the information, no inclination is given to any of the qualities and each characteristic is considered similarly to find the examples. The execution assessment of the proposed demonstrate is done on UCSD information mining challenge 2009 dataset (mysterious also, imbalanced) and it is discovered that the proposed display has high extortion location rate, adjusted arrangement rate, Matthews connection coefficient, and less false alert rate than other best in class classifiers.

This paper proposed an extortion identification show whose execution is assessed with an anonymized dataset and it is discovered that the proposed display functions admirably with this sort of information since it is autonomous of quality esteems. The second component of the proposed show is its capacity to deal with class imbalance. This is joined in the model by making two separate example databases for misrepresentation and legitimate exchanges. Both client and fake practices are observed to change progressively finished a more drawn out timeframe.

Yigit Kultur et.al [7] described about the visa extortion costs the saving money segment billions of dollars consistently, diminishing the misfortunes acquiring from visa extortion is an essential driver for the part and end-clients. Manage based extortion location instruments have been generally utilized as a piece of charge card frameworks. Tenets of such devices are dictated by human extortion specialists. In any case, specialists for the most part disregard cardholder-particular spending conduct. In this paper, we concentrate on breaking down the cardholder spending conduct and propose a novel cardholder conduct show for identifying credit Card misrepresentation.

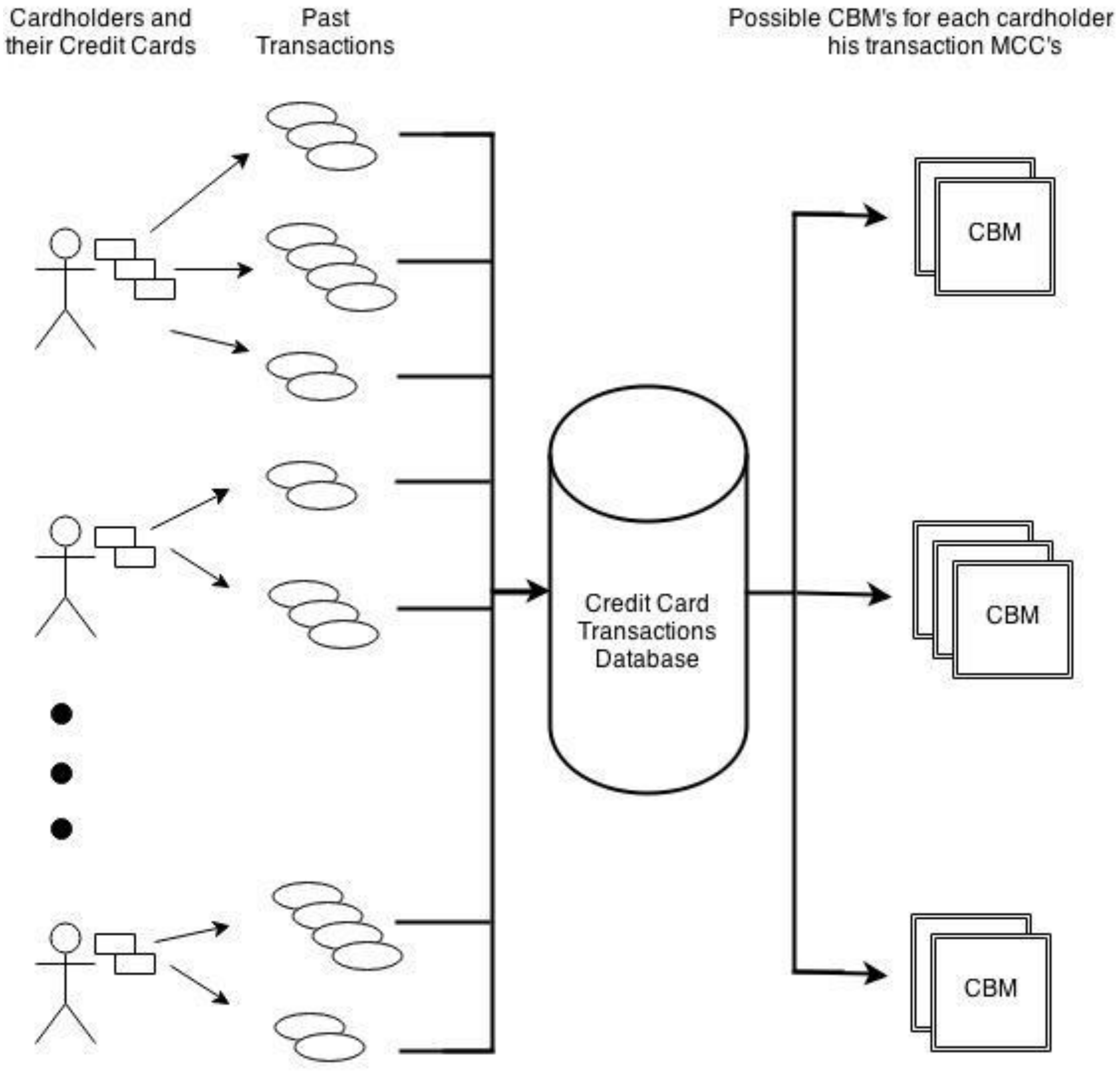

Cardholders and their Credit Cards
Past Transactions
Possible CBM's for each cardholder his transaction MCC's
Credit Card Transactions Database
CBM
CBM
CBM

Figure 2.1: Cardholder behaviour model (CBM) for credit card fraud detection

In this paper, the authors have experimentally demonstrated the impact of the proposed show on misrepresentation identification execution. The common sense effect of CBM is to give a steady extortion location device which will cooperate with the current run based devices. CBM identifies around 43 for every penny of the fakes by displaying the conduct of the cardholders. In future research, the author's expect to rehash the investigations with datasets from other driving banks in turkey to limit the danger to outside legitimacy.

Andrea Dal Pozzolo et.al [7] distinguishes about the fake transaction in master card exchanges is maybe outstanding amongst other test beds for computational knowledge calculations. Actually, this issue includes various significant challenges, in particular: idea float (clients propensities develop and fraudsters change their techniques after some time), class irregularity (authentic exchanges far dwarf fakes) and check inactivity (just a little arrangement of exchanges are opportune checked by examiners). In any case, most by far of learning calculations that have been proposed for misrepresentation recognition depends on presumptions that barely hold in a true misrepresentation discovery framework (FDS).

Here author dissect in detail this present reality working states of FDS and give a formal depiction of the verbalized order issue included. Specifically, author has depicted the ready input association, which are the instrument giving late regulated specimens to prepare the classifier. Author additionally assert that, conversely with conventional execution measures utilized as a part of the writing, in a true FDS, the exactness of the detailed alarms is likely the most important one, since specialists can check just couple of cautions.

Author investigated on two huge datasets of genuine exchanges demonstrate that, keeping in mind the end goal to get exact alarms, it is obligatory to dole out bigger significance to criticisms amid the learning issue. As anyone might expect, criticisms assume a focal part in the proposed learning technique.

Table 2.1: Comparison of Credit Card Fraud Detection Technique

Index	Authors, Publisher & Year	Problem Addressed	Technologies Used	Remarks
1	Kulkarni et. al. [1], Springer 2016.	Credit card fraud determination on the German credit card data	Logistic Regression for classification of the unbalanced metrics in the heterogeneous environment	Remarkable performance in the terms of accuracy, which is slightly lower from the real-time requirements. Also the MAE must be reduced further to create a dependable system.
2	Bahnsen et. al. [2], Elsevier 2016.	Feature engineering for the credit card fraud detection.	Periodic behaviour based feature presentation along with Mises distribution.	The savings can be further increased by engineer the prominent features.
3	Dal Pozzolo et. al. [3] Elsevier 2014.	Analyse the various questions related to the crucial issues in credit card fraud detection.	Discussed unbalanced features, assessment and non-stationary features.	Single transaction in the history can't play the role of vital metric for fraud detection.
4	Halvaiee et. al. [4] Elsevier 2014.	Credit card fraud detection using artificial	CC Fraud using artificial immune	The false positive rate is recorded between 0.017 and 0.035. The false

		immune systems	inspired system (AIRS) with distance and scoring aware metric similarity evaluation.	positive rate can be further reduced by using the deep feature engineering methods to decrease the losses.
5	Van Vlasselaer et. al. [5] Elsevier 2015.	Network based Extensions for the CC Fraud detection	Anomaly prevention using advanced transaction exploration (APATE)	The AUC percentage 0.986, which can be further improved by balancing the unbalanced metrics.

CHAPTER 3

SCOPE OF THE STUDY

The credit card frauds are on the rise with the rise in the number of credit card holders. The credit card frauds takes place with the non-awareness of the holders or IT based expertise of the hackers, which as a result accounts for the loss of heavier amounts every year. To commit the credit card fraud, the hackers utilize the various methods from the social engineering, point of sales hacking, phishing attacks, etc. to steal the information related to the credit cards. The credit card fraud detection model in existing scheme is utilizing the unbalanced metrics and practices to balance them in order to produce the decision, whether the credit card fraud has taken place or not.

The feature scaling has been found missing from the existing model design, which account for the dilution of the data column dominance. The feature scaling methods can further improve the overall accuracy. The mean absolute error (MAE) has been recorded with the value of 74.83%, which is considerably very high and must be curbed or neutralized in order to achieve the higher accuracy. The presence of the statistical type 1 and 2 errors (i.e. False Positive & False Negative) in the results has been indicated by the relative absolute error (RAE), which has been recorded at 0.406 and have resulted the lower accuracy based results of 96.6243%.

In order to eliminate the problems in the existing model, the proposed model will be designed with the unbalanced metric normalization methods, where the combination of averages and floating averages (such as Mean, Median, Standard Deviation, Floating Mean, Variance, Covariance etc.) can be utilized to create the state-of-art system in order to minimize the feature unbalance in the feature matrix.

In addition to averaging factor based feature description, the flexible and robust feature scaling practices can be utilized, which may vary from column to column in the given data according to its volatility and overall variance, to precise the features in order to create the high accuracy based credit card fraud detection model. The model with best feature selection with probabilistic classification (BFS-PC) for the purpose of credit card fraud detection would be deployed with certain improvements or enhancements during the proposed model implementation.

The best feature selection method will incorporate the selection of the features on the basis of their compatibility, which can be measured with the column or feature variance. The probabilistic classification algorithm involves the probability based matching between the training and testing data, which is decided with the maximum likeliness or similarity between the entries of test and train data.

Chapter 4

OBJECTIVE OF THE STUDY

The main objectives of the study are as following:

- To study the various credit card fraud detection techniques in order to recognize their advantages and shortcomings.
- To design an improved credit card fraud detection scheme using the data classification.
- To simulate and compare the results of the proposed scheme with the existing techniques of credit card fraud detection.

Chapter 5

RESEARCH METHODOLOGY

At first stage, a detailed literature study would be conducted on the credit card fraud detection algorithm and data classification methods; and to know their advantages and disadvantages. Literature study will lead us towards refining the structure of the proposed security solution design to overcome the shortcomings of the existing schemes, while keeping their advantages intact in order to build a robust system. Afterwards, the proposed solution will be implemented in PYTHON simulator with all essential input and output parameters. Then the implementation will undergo a thorough performance analysis and detailed comparison with the existing models.

The input data acquisition is done on the historical data of the credit card for last 1 to 10 years. The historical data of credit card spending contains the readings of the spending based listings in the day to day interval analysis, which contains the starting and closing value of each day accounted in the historical data. During the implementation, the proposed model would be designed using the ensemble of the regression models with squared distance based classification for the purpose of historical data processing, which is generally utilized for the long term prediction.

After the classification, the intensity of the individual cards is inquired and calculated, which prepares and classifies the credit card trend and spending based anomalies. The regression models are used to perform the operation on the given data stream obtained from the credit card company for the detection of the credit card frauds by analysing the spending behaviour of the customers. The structural anomaly modelling is used to detect the credit card frauds with flexibility and variability relationship and payment based relation building, which predicts the spending based anomalies, which leads towards the credit card fraud based decisions.

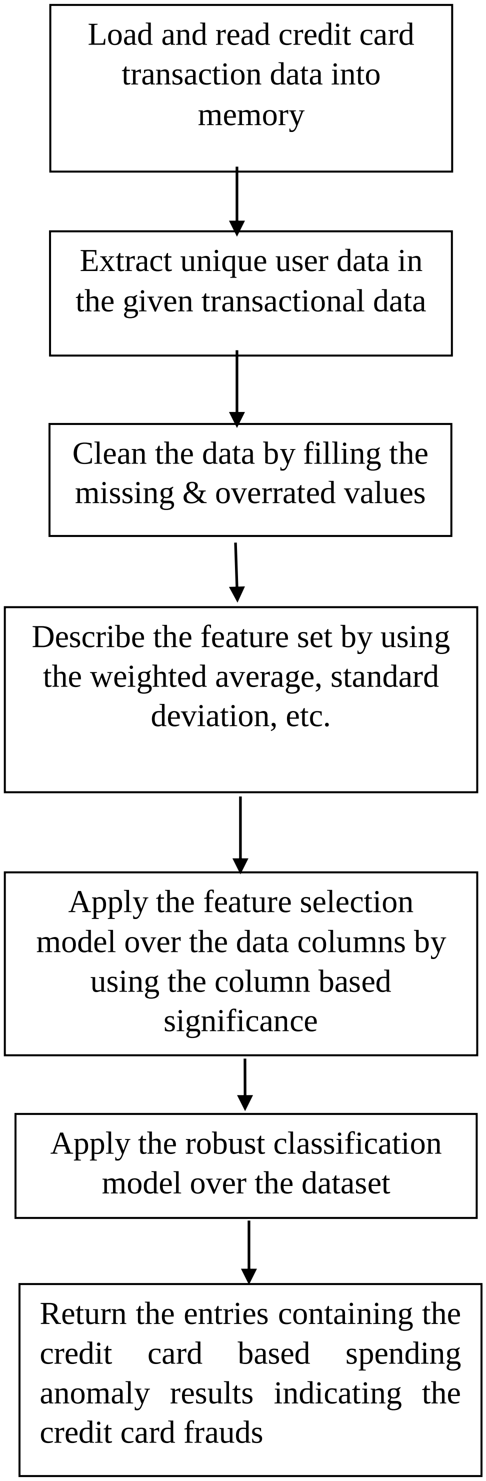

Load and read credit card transaction data into memory
Extract unique user data in the given transactional data
Clean the data by filling the missing & overrated values
Describe the feature set by using the weighted average, standard deviation, etc.
Apply the feature selection model over the data columns by using the column based significance
Apply the robust classification model over the dataset
Return the entries containing the credit card based spending anomaly results indicating the credit card frauds

Figure 5.1: Methodology

5.1 HARDWARE AND SOFTWARE REQUIRED

- **Hardware:**

a) A simple PC (With Wireless NIC embedded)

b) 100 GB storage

c) 2 GB Ram

d) Multi Core Processor

e) Wireless LAN, Wi – Fi

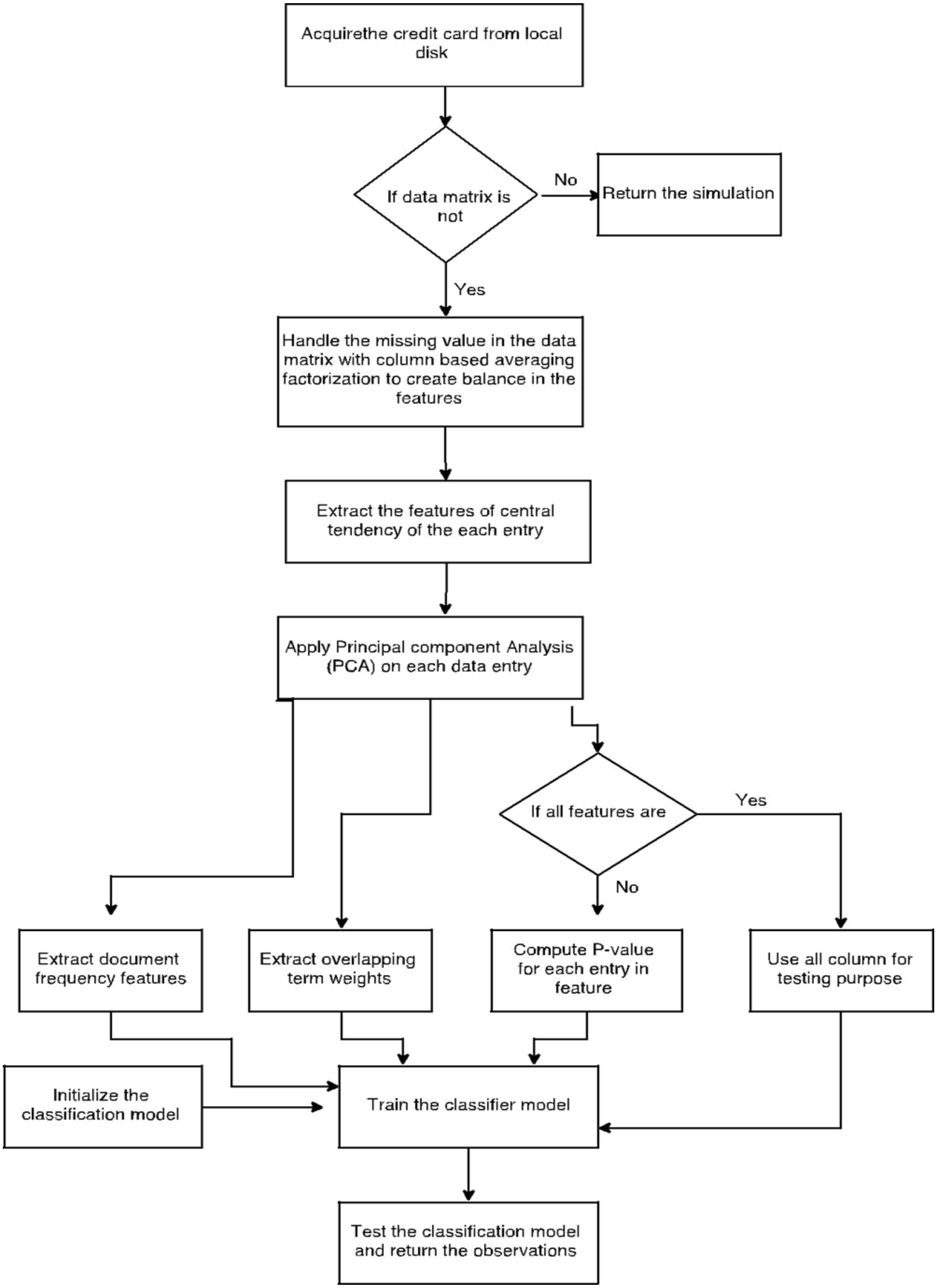
- **Software:**

a) OS: Windows

b) Integrated Development Environment (IDE): Anaconda with Spyder (Python version 2.7 or 3.6)

c) Programming Language: PYTHON.

5.2 FLOW CHART


Acquirethe credit card from local disk
If data matrix is not
No
Return the simulation
Yes
Handle the missing value in the data matrix with column based averaging factorization to create balance in the features
Extract the features of central tendency of the each entry
Apply Principal component Analysis (PCA) on each data entry
If all features are
Yes
No
Extract document frequency features
Extract overlapping term weights
Compute P-value for each entry in feature
Use all column for testing purpose
Initialize the classification model
Train the classifier model
Test the classification model and return the observations

Chapter 6

EXPECTED OUTCOMES

- The unbalanced data matrices make the credit card fraud detection task tougher by adding the volatility over the given set of data. The unbalanced data distribution requires the balanced feature description in order to normalize the unexpectedly high or low frequencies, which can further improve the performance of the credit card fraud detection models.
- The feature scaling has been performed in the base model at par according to the proposed framework (Described in section 3.1), which is also a reason behind the higher errors, specifically mean absolute error (MAE), which accounts for 74.83% and considerably very high.
- The relative absolute error has been recorded at 0.406, and shows the presence of false positive and false negative cases, which eventually reduces the overall accuracy to 96.6243%. The accuracy must be improved in order to realize the state-of-art model for the credit card fraud detection.

Chapter 7

SUMMARY AND CONCLUSIONS

The credit card fraud detection methods have gained the popularity in the past decade with the evolution of the statistical models. These models are used to automate the process of pattern recognition, which takes comparatively lesser time and can handle a large number of transactions per day.

The statistical model based upon the binary classification such as support vector machine (SVM), stochastic gradient descent (SGD), Adaboost, etc. to improve the accuracy of the credit card fraudulent pattern detection. The malicious patterns are also known as outlier or anomaly, which must be detected correctly in order to minimize the bank losses caused by fraudulent transactions. The performance of the proposed model would be evaluated using the precision, recall, F1-measure and accuracy based parameters.

By using the supervised learning techniques we are trying to improve the accuracy of the credit card fraud detection based on the data classification method. As the world is becoming digital day by day as a result there is increasing rate of fraud transactions.

REFERENCES

[1] Kulkarni, Pallavi, and Roshani Ade. "Logistic Regression Learning Model for Handling Concept Drift with Unbalanced Data in Credit Card Fraud Detection System." In *Proceedings of the Second International Conference on Computer and Communication Technologies*, pp. 681-689. Springer India, 2016.

[2] Bahnsen, Alejandro Correa, Djamila Aouada, Aleksandar Stojanovic, and Björn Ottersten. "Feature engineering strategies for credit card fraud detection." *Expert Systems With Applications* 51 (2016): 134-142.

[3] Dal Pozzolo, Andrea, Olivier Caelen, Yann-Ael Le Borgne, Serge Waterschoot, and Gianluca Bontempi. "Learned lessons in credit card fraud detection from a practitioner perspective." *Expert systems with applications* 41, no. 10 (2014): 4915-4928.

[4] Halvaiee, Neda Soltani, and Mohammad Kazem Akbari. "A novel model for credit card fraud detection using Artificial Immune Systems." *Applied Soft Computing* 24 (2014): 40-49.

[5] Van Vlasselaer, Véronique, Cristián Bravo, Olivier Caelen, Tina Eliassi-Rad, Leman Akoglu, Monique Snoeck, and Bart Baesens. "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions." *Decision Support Systems* 75 (2015): 38-48.

[6] Prakash, A., and C. Chandrasekar. "An optimized multiple semi-hidden markov model for credit card fraud detection." *Indian Journal of Science and Technology* 8, no. 2 (2015): 165-171.

[7] Bahnsen, Alejandro Correa, Aleksandar Stojanovic, Djamila Aouada, and Björn Ottersten. "Improving credit card fraud detection with calibrated probabilities." In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pp. 677-685. Society for Industrial and Applied Mathematics, 2014

[8] Zareapoor, Masoumeh, and Pourya Shamsolmoali. "Application of credit card fraud detection: Based on bagging ensemble classifier." *Procedia Computer Science* 48 (2015): 679-685.

[9] Seeja, K. R., and Masoumeh Zareapoor. "FraudMiner: a novel credit card fraud detection model based on frequent itemset mining." *The Scientific World Journal* (2014).