

# **CREDIT CARD FRUAD DETECTION USING MACHINE LEARNING**

## **A PROJECT REPORT**

*Submitted by*

**GAYATHRI. J**

**[513220104701]**

**LAVANYA. V**

**[513220104702]**

**LAVANYA. S**

**[513220104004]**

**SOWMIYA. R**

**[513220104315]**

*In partial fulfillment for the award of the degree  
of*

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**



**THIRUMALAI ENGINEERING COLLEGE, KANCHIPURAM**

**ANNA UNIVERSITY: CHENNAI – 600 025**

**MAY 2024**

**ANNA UNIVERSITY: CHENNAI 600 025**

**BONAFIDE CERTIFICATE**



Certificate that this project report titled **“CREDIT CARD FRUAD DETECTION USING MACHINE LEARNING”** is the bonafide work of **“GAYATHRI. J [513220104701], LAVANYA. V [513220104702], LAVANYA. S [513220104004], SOWMIYA. R [513220104315]”** who Carried out the project work under my supervision.

**SIGNATURE OF HOD**

**V. VIJAYABHASKAR M.C.A., M.Tech.,**

**HEAD OF THE DEPARTMENT,**

Associate Professor,

Department of CSE,

Thirumalai Engineering College,

Kanchipuram – 631 551.

**SIGNATURE OF SUPERVISOR**

**S. HEMALATHA M.E.,**

**SUPERVISOR,**

Assistant Professor,

Department of CSE,

Thirumalai Engineering College,

Kanchipuram – 631 551.

Submitted for the Project Viva Voce held on \_\_\_\_\_

INTERNAL EXAMINER

EXTERNAL EXAMINER

## ACKNOWLEDGEMENT

I profoundly thank our **Chairman and trust members of Kanchipuram Educational Trust** for providing adequate facilities.

I would like to express my hearty thanks to our respectable Principal. Incharge **Mr.T.MohanRaj M.Tech.**, for allowing us to have the extensive use of our colleges facilities to our colleges facilities to have precious advice regarding the project.

I extend our thanks to Associate Professor **Mr.V.VIJAYABHASKAR M.C.A., M.Tech., Head of the Department, Information Technology** for this precious advice regarding the project.

I would like to express my deep and unbounded gratefulness to my project Guide **Mrs.S.HEMALATHA M.E.**, Department of Information Technology, for his valuable guidance and encouragement throughout the project. He has been a constant source of inspiration and has provided the precious suggestion throughout this project.

I thank all facilities and supporting staff for the help they extended in completing this project. I also express my sincere thanks to our parents, and all my friends for their continuous support.

## **TABLE OF CONTENTS**

<b>CHAPTER NO</b>	<b>LIST OF CONTENT</b>	<b>PAGE NO</b>
	<b>ABSTRACT</b>	<b>I</b>
	<b>LIST OF ABBREVIATION</b>	<b>II</b>
	<b>LIST OF FIGURES</b>	<b>III</b>
	<b>LIST OF TABLES</b>	<b>V</b>
<b>1</b>	<b>INTRODUCTION</b>	
	1.1 GENERAL	1
	1.2 OBJECTIVES	2
	1.3 EXISTING SYSTEM	3
	1.3.1 DISADVANTAGES OF EXISTING SYSTEM	4
	1.4 PURPOSE	5
	1.5 SYSTEM PROPOSED	6
<b>2</b>	<b>LITERATURE SURVEY</b>	
	2.1.1 PAPER 1	7
	2.1.2 PAPER 2	8
	2.1.3 PAPER 3	9
	2.1.4 PAPER 4	10
	2.2 CREDIT CARD FRUDULENT DETECTION	12
	2.3 DATA SAMPLING	12
	2.4 CREDIT CARD FRUDULENT DETECTION USING HIDDEN MARKOV MODEL	13
		6
		7

2.5 CREDIT CARD FRUDULENT DETECTION	13
USING DECISION TREE INDUCTION	
ALGORITHM	

3	<b>CREDIT CARD FRUDULENT DETECTION</b>	
	<b>SYSTEM</b>	
	3.1 GENERAL	15
	3.2 PROBLEM DEFINITION OF CREDIT CARD	16
	FRUDULENT	
	3.3 BLOCK DIAGRAM	17
	3.4 METHODOLOGY	17
	3.4.1 WHAT ARE ANOMOLIES	18
	3.4.2 ANOMOLIES DETECTION	19
	3.4.3 NOISE REMOVAL	19
	3.5 ANOMOLY DETECTION TECHNIQUES	19
	3.5.1 SIMPLE STATISTICAL METHODS	19
	3.5.2 CHALLENGES WITH SIMPLE	19
	STATISTICAL METHODS	
	3.6 CREDIT CARD FRUDULENT DETECTION	20
	SYSTEMS	
	3.7 FUNCTION ALITIES	20
	3.8 ACCURACY	21
	3.9 OBSERVATION	21
	3.10 MODEL PREDICTION	22
	3.10.1 ISOLATION FOREST ALGORITHM	23
	3.10.2 WORKING PRINCIPAL OF ISOLATION	23
	RANDOM FOREST	
	3.10.3 LOCAL OUTLIER FACTOR	23

	3.10.4 OBSERVATION	24
4	<b>INTRODUCTION TO MACHINE LEARNING</b>	
	4.1 GENERAL	25
	4.2 OVERVIEW OF MACHINE LEARNING	26
	4.3 MACHINE LEARNING BASED APPROCHES	27
	4.3.1 DENSITY BASED DETECTION OF ANOMALY	27
	4.3.2 CLUSTERING DETECTION OF ANOMALY	28
	4.3.3 SVM BASED DETECTION OF ANOMALY	28
	4.4 DATASET	29
	4.4.1 DATASET DETAILS	30
5	<b>SYSTEM SPECIFICATION</b>	
	5.1 GENERAL	31
	5.2 HARDWARE REQUIREMENTS	31
	5.3 SOFTWARE REQUIREMENTS	32
	5.4 SOFTWARE USED	32
6	<b>DESIGN ENGINEERING</b>	
	6.1 GENERAL	33
	6.2 ACTIVITY DIAGRAM	34
	6.3 USE CASE DIAGRAM	35
	6.4 SEQUENCE DIAGRAM	36
	6.5 CLASS DIAGRAM	37

	6.6 DATA FLOW DIAGRAM	38
	6.7 COMPONENT DIAGRAM	39
	6.8 DEPLOYMENT DIAGRAM	40
7	<b>IMPLEMENTATION</b>	
	7.1 GENERAL	41
	7.2PROCEDURE FOLLOWED DURING IMPLEMENTATION	41
	7.2.1 DATASET DESIGN	42
	7.2.2 DATA DESCRIBE	43
	7.2.3 PREPROCESSING	44
	7.2.4 FIND FRAUD	46
	7.2.5 HEATMAP	46
	7.2.6 PREDICTION	47
8	<b>SOFTWARE TESTING</b>	
	8.1 GENERAL	49
	8.2 TESTING	49
9	<b>CONCLUSION</b>	52
10	<b>ALGORITHM</b>	
	10.1 LOGISTIC ALGORITHM	53
	10.1.1 ADVANTAGES OF LOGISTICS ALGORITHM	54
	10.2 RANDOM FOREST CLASSIFIER	54
11	<b>IMPLEMENTATION</b>	56

12	<b>APPLICATION AND FUTURE WORK</b>	
	12.1APPLICATION	58
	12.2 FUTURE ENHANCEMENTS	59
13	<b>SOURCE CODING</b>	60
14	<b>OUTPUT</b>	70
15	<b>CONCLUSION AND FUTURE ENHANCEMENT</b>	74
16	<b>REFERENCES</b>	77



## ABSTRACT

Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. Increase in fraud rates, researchers started using different machine learning methods to detect and analyse frauds in online transactions.

The main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. Where cardholders are clustered into different groups based on their transaction amount.

Then using sliding window strategy, to aggregate the transaction made by the cardholders from different groups so that the behavioural pattern of the groups can be extracted respectively. Later different classifier are trained over the groups separately. And then the classifier with better rating score can be chosen to be one of the best methods to predict frauds.

Thus, followed by a feedback mechanism to solve the problem of concept drift. In this paper, we worked with European credit card fraud dataset.

**Keywords:** Card-Not-Present frauds, Card-Present-Frauds, Concept Drift.

## **LIST OF ABBREVIATIONS**

### **ACRONYM**

### **ABBREVIATIONS**

WHO	WORLD HEALTH ORGANIZATION
NLP	NATURAL LANGUAGE PROCESS
ML	MACHINE LEARNING
EDA	EXPLORATORY DATA ANALYSIS
CSV	COMMA SEPERATE VALUE
KNN	K-NEAREST NEIGHBOR
ROC	RECEIVER OPERATING CHARACTER
API	APPLICATION PROGRAMMABLE INTERFACE
NOSQL	NOT ONLY SQL
VOC	VARIENCES OF CONCERN
SVM	SUPPORT VECTOR MECHINE
BDV	BIG DATA VISUALIZATION

## LIST OF FIGURES

FIGURE NO	FIGURES	PAGE NO
3.3.1	BLOCK DIAGRAM	17
3.2.1	ANOMOLY DETECTION	18
6.2.1	ACTIVITY DIAGRAM	34
6.3.1	USE CASE DIAGRAM	36
6.4.1	SEQUENCE DIAGRAM	37
6.5.1	CLASS DIAGRAM	38
6.6.1	DATA FLOW DIAGRAM	39
6.7.1	COMONENT DIAGRAM	39
6.8.1	DEPLOYEMENT DIAGRAM	40
7.1.1	DATASET DIAGRAM	43
7.2.1	DATA DESCRIBE	43
7.3.1	HISTGRAM	45
7.3.2	FRUAD DIAGRAM	46
7.3.3	HEAT MAP DIAGRAM	47
7.3.4	ACCURACY DIAGRAM	48
10.2.1	RFC DIAGRAM	55
14.1.1	FRONT END FRAMEWORK	70
14.2.1	DATA PREPROCESSING	71
14.3.1	ACCURACY	72
14.4.1	HEAT MAP	73

## **LIST OF TABLES**

<b>TABLE</b>	<b>TABLES</b>	<b>PAGE</b>
<b>NO</b>		<b>NO</b>
8.2.1	TESTING RESULTS	49

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 GENERAL**

Nowadays as we can see that there is a huge increase online payment and the payment is mostly done with the help of credit cards. It becomes a big problem for marketing company to overcome with the credit card fraudulent activities. Fraudulent can be done in many ways such as tax return in any other account, taking loans with wrong information etc. Therefore, we need an efficient fraudulent detection model to minimize fraudulent activity and to minimize their losses. There are a huge number of new techniques which provide different algorithms which help in detecting number of credit card fraudulent activity. Basic understanding of these algorithms will help us in making a significant credit card fraudulent detection model. This paper helps us in finding doubtful credit card transaction by proposing a machine learning algorithms. Credit Card Fraudulent detection comes under machine learning, and the objective is to reduce such type of fraudulent activity. This type of fraud is happening from past, and till now not much research has done here in this particular area. The types of credit fraud in transactions are bankruptcy fraud, behavioral fraud, counterfeit fraud, application fraud. There are experiments done before on credit card fraudulent activity on basis of meta-learning. There is certain limit of meta-learning. There are two features which is introduced here in our report is True Positive and False alarm. Both these features play an important role in catching fraudulent because the rate of determining fraudulent behavior is quick. For the better performance of model, we need a better classifier. Different classifier can be combined together with help of meta-learning.

## **1.2 OBJECTIVES**

To run a suitable business, vendors need to make a profit, which can be calculated by subtracting the cost of doing business from the total sell price. Therefore, fraudulent become a business's tolerance among online payment, among financing company, gross margin is calculated by (sell price - cost of goods sold). The lower the margin, there will be low risk for fraudulent payment. In practice, whenever fraudulent occurs, the cardholder have to complain to the financing company and the debit from card is usually cancelled, which means there is a loss for either cardholder's bank or the finance company. Fraudulent turns as a financial risk to the financial company and the cardholder's bank. To overcome with fraudulent, fraudulent detection techniques should be used. The main objective is to prevent the customer from fraud because if this kind of things keep happening then people will not show there interest in taking credit card and using there facility which is given by the banks and other financial company.

Therefore, it's become an essential thing nowadays. People should also takes care of their personal information by keeping it to the limited source. The fraudulent activity start with the leaking of the someone personal information like credit number, one time password, registered mobile number and many more. The sharing of someone personal information should be reduced because fraudulent activity begin with the help of someone personal information like credit card number and many more.

### 1.3 EXISTING SYSTEM

The previous detecting technique takes a long time to catch fraud which is basically depend on the database, not that much accurate and not give the result in-time. After that algorithm which is used for the detection of credit card fraudulent is generally on basis of analysis, fraudulent detection based on credit card transaction made by cardholder and the credit rate for cardholders.

There are certain limits of meta-learning. There are two features which is introduced here in our report is True Positive and False alarm. Both these features play an important role in catching fraudulent because the rate of determining fraudulent behavior is quick. For the better performance of model, we need a better classifier. Different classifier can be combined together with help of meta-learning.

Previously attempts have been made to work out Credit Card Fraud Detection system using SVM (Select Vector Machine). SVM makes use of hyperplane to classify the data points in a collection. A good hyperplane associates greater number of data points within its margin <sup>[2]</sup>.

This is not efficient for a large amount of data sets. As, in large amount of data sets there is a probability of redundant data which will take more time to process.

Therefore, it usually delayed in calculating the fraud or there might be probability to not calculate in time.

### **1.3.1 DISADVANTAGES OF EXISTING SYSTEM**

- In case of fraud there is a high amount losses and thus because of this loss, card limit should be reduced.
- The fraudulent should be detected in real time and omission in false transactions is mandatory.
- Reasons of fraudulent should be identified from data available.
- System should be capable in identifying the trend of fraud transaction.
- Machine learning algorithm work only for huge sets of data.
- For smaller amount of data the results may be not accurate.
- It takes significant amount of data for machine.



## **1.4 PURPOSE**

The purpose of the "Credit Card Default Prediction System" project is to create a software that helps in preventing credit card frauds. This project has several key objectives:

- Enhanced Security:** The primary purpose is to enhance the security of credit card transactions, protecting both financial institutions and cardholders from fraudulent activities.
- Minimize Financial Losses:** The project aims to minimize financial losses resulting from credit card fraud, reducing the burden on financial institutions and individual cardholders.
- Real-time Detection:** The project seeks to develop a system capable of real-time detection of suspicious and potentially fraudulent transactions, enabling immediate action to prevent further fraud.
- Machine Learning Models:** Utilize machine learning algorithms to create advanced models that can effectively distinguish between legitimate and fraudulent credit card activities.
- Transparent Reporting:** Implement a transparent reporting mechanism to notify relevant stakeholders, including cardholders and financial institutions, of potential fraudulent transactions, fostering trust and prompt action.
- Data-Driven Insights:** Leverage data analytics to provide valuable insights into fraud patterns, enabling proactive measures to combat emerging fraud tactics.
- Compliance and Ethical Standards:** Ensure that fraud detection practices adhere to ethical standards and legal regulations, balancing security with individual privacy rights.
- Adaptability:** Design the system to be adaptable and responsive to evolving fraud methods and tactics in the ever-changing landscape of financial transactions.
- Collaboration:** Foster collaboration between financial institutions, merchants, and technology providers to collectively combat credit card fraud and promote a secure financial environment.

## 1.5 THE SYSTEM PROPOSED

- In this model we overcome with the issues in a significant way. Using Isolation random forest and local outlier factor algorithm we can detect the fraud in actual time and find out the way to minimize the fraud to produces an optimized result so that it will perform a better prediction.

On the basis of customer's behavior, we can detect fraudulent.

- We have used logistic regression and random forest. We can get more accuracy like 0.99 etc....
- We are taking the dataset with help of simple GUI from our local directory where we downloaded the dataset.
- With the help of random forest algorithm and local outlier factor we are finding the data point which is different from its neighbor and can be a fraudulent transaction with its outlier behavior.
- We have two classification class which is named as class 0 and class 1.
- Random forest is the [powerful machine learning algorithm is commonly used is credit card fraud detection due its ability to handle large large dataset is high dimension and imbalance classes.
- User can safety use is credit card for online transaction.
- Adder layer of security.
- Reduction in number of fraud transaction.

## **CHAPTER 2**

### **LITERATURE SURVEY**

**1. Delamaire. L. Abdou, HAH and Pointon. J,”Credit card fraud and detection techniques”, Banks and Bank Systems, Volume 4, Issue 2, 2009.**

Data mining concerns the extraction of implicit knowledge, data relationship or other patterns not explicitly stored in the large amount of data. Fraud mining in large amount of data is one of the powerful sources of high-level semantics. If these fraudulent transactions could be identified, detected and recognized automatically, they would be a valuable source of high-level semantics for indexing and retrieval. This thesis developed to analyze, detect and recognize the fraudulent transactions and the system is based on efficient clustering and classification methods such as apriority and support vector machine respectively. The result shows that the proposed method gives better results which helps to obtain high fraud coverage combined with a low false alarm rate than the existing Hidden Markov Model.

In our paper we referred to various papers for improving the performance of routing, reduce delay of information, reduce packet loss rate, reduce link failure, to improve packet delivery rate, to reduce energy consumption. There are a huge number of new techniques which provide different algorithms which help in detecting number of credit card fraudulent activity. Basic understanding of these algorithms will help us in making a significant credit card fraudulent detection model. This paper helps us in finding doubtful credit card transaction

by proposing a machine learning algorithms. There are two features which is introduced here in our report is True Positive and False alarm. Both these features plays an important role in catching fraudulent because the rate of determining fraudulent behavior is quick. As per today's Network plays an important role therefore it is mandatory for our models to be up to date to perform better detection capabilities. Whenever new fraudulent activity are detected then our model should be that much better to perform real time analysis. Other than traditional machine learning methods Fraudulent Detection System has been achieved through using Neural Networks <sup>[5]</sup>. To prevent personal information has become a huge task for financial company because there are a lot of attack on the system to steal someone personal information to perform fraudulent. Our model has two essential feature which will help in finding abnormal behavior in form of charts for different column such as time, amount etc.

**2. Suman, Nitin, “Review Paper on Credit Card Fraud Detection”,  
International Journal of Computer Trends and Technology (IJCTT) –  
volume 4 Issue 7–July 2013.**

Review Paper on Credit Card Fraud Detection 1 Suman Research Scholar, GJUS&T Hisar HCE Sonapat 2 Nutan Mtech.CSE, and HCE Sonapat  
Abstract Due to the theatrical increase of fraud which results in loss of dollars worldwide each year, several modern techniques in detecting fraud are persistently evolved and applied to many business fields. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid undesirable behavior. Undesirable behavior is a broad term

including delinquency, fraud, intrusion, and account defaulting. This paper presents a survey of current techniques used in credit card fraud detection and telecommunication fraud. The goal of this paper is to provide a comprehensive review of different techniques to detect fraud. Keywords: Fraud detection, data mining, support vector machine, anomalies.

**Introduction** Credit card fraud can be defined as unauthorized account activity by a person for which the account was not intended. Operationally, this is an event for which action can be taken to stop the abuse in progress and incorporate risk management practices to protect against similar actions in the future. In simple terms, Credit Card Fraud is defined as when an individual uses another individual's credit card for personal reasons while the owner of the card and the card issuer are not aware of the fact that the card is being used. And the persons using the card has not at all having the connection with the cardholder or the issuer and has no intention of making the repayments for the purchase they done. Fraud detection involves identifying Fraud as quickly as possible once it has been perpetrated. Fraud detection methods are continuously developed to defend criminals in adapting to their strategies. The development of new fraud detection methods is made more difficult due to the severe limitation of the exchange of ideas in fraud detection. Data sets are not made available and results are often not disclosed to the public. The fraud cases have to be detected from the available huge data sets such as the logged data and user behavior. At present, fraud detection has been implemented by a number of methods such as data mining, statistics, and artificial intelligence. Fraud is discovered from anomalies in data and patterns. The different types of methods for committing credit card frauds are described below. Types of Frauds: Various types of frauds in this

paper include credit card frauds, telecommunication frauds, and computer intrusions, Bankruptcy fraud, Theft ISSN: Page 2206.

**3.Renu, Suman, “Analysis on Credit Card Fraud Detection Methods”, International Journal of Computer Trends and Technology (IJCTT) – volume 8 number 1 – Feb 2014.**

The advancement of new technologies and the fast growing of technological development have generated new possibilities as well as imposing new challenges. Fraud, the biggest challenges for business and organization, emerge with new technologies to take new and distinctive forms that are hidden and tougher to identify than the conventional forms of this crime. Credit card frauds also grow up along with growing in technology. It also noticed that financial fraud is extremely growing in the global communication improvement. It is being admitted every year that the loss because of this types of fraudulent activities is billions of dollars. These activities are performed so gracefully that it look similar to original transactions. Simply using of pattern matching technique and simple method really not useful for detecting these fraudulent activities. A well planned and systematic method has become need for all business and organization to minimizing chaos and carry out in place. Several technique has been evolved based on Artificial intelligence, Machine learning, Data mining, Genetic programming Fuzzy logic etc... For detecting credit card fraudulent activities. Besides this technique, K-Nearest Neighbor algorithm and outlier detection methods are implemented to optimize the best solution for the fraud detection problem. These techniques proved to minimize the false alarm rates and increase the fraud detection rate.

**4. V. Bhusari S. Patil, “Study of Hidden Markov Model in Credit Card Fraudulent Detection”, International Journal of Computer Applications (0975 – 8887) Volume 20– No.5, April 2011.**

He results confirmed that the proposed model outperforms the rule-based algorithms in terms of false positive rate. In the same context of false positive rate, in 2011 Bhusari V. et al. used Hidden Markov Model in order to detect credit card fraud during transactions. Their experiment confirmed that HMM model helps to obtain a high fraud reporting combined with a low false positive. ...

Their experiment confirmed that HMM model helps to obtain a high fraud reporting combined with a low false positive. HMM model represents a great value solution for addressing detection of fraud transaction through credit card. Also, Delio Panaro et al. (2015) [12] proposed a two layer statistical classifier for sensitive, highly skewed and massive data sets to detect fraud. ...

E-commerce fraud 9 [8], [11], [12], [34], [35], [47], [51], [53], [57] We find 19 machine-learning techniques mentioned in these articles analyzed with different numbers per each article, from 1 to maximum 5. Also, there are 4 techniques (Artificial Neural Network, Decision Tree, Genetic algorithm, and Support Vector Machine) which appear in more 10 articles. On the other hand, there are 4 techniques (Expert system, Gradient Descendent, K-means, and Scatter Search) mentioned in only one article.

## **2.2 CREDIT CARD FRAUDULENT DETECTION**

We publish a Credit Card fraudulent detection model whose performance is evaluated on basis of anonymized data sets and found that detection model performance is good for this dataset. This is incorporated that this model creates two separate patterns for databases, one for fraud and other for legal transactions. The fraudulent detection model should be more accurate in order to detect the changing behavior of consumer and his behavior. We can predict this fraudulent by running our model after every fixed amount of transaction or after a fixed interval of time. AI provides procedure for various types of calculations which can be performed independently. If there is any outlier value in our dataset, then our model can detect it. Outlier value means the value which deviates by a long margin from their neighbor can perform abnormal behavior. That outlier behavior is the fraudulent transaction in dataset. We have also reduced redundancy of datasets by removing some of the redundant data from our dataset. Because our main aim is achieving the real time analysis and for that we need to reduce the datasets so that we can speed up our algorithm performance.

## **2.3 DATA SAMPLING**

Since, Random forest algorithm is a machine learning algorithm therefore we need trained dataset to perform our mechanism. These trained datasets are then loaded to the main memory of the system. Our dataset has almost 300,000 value so it's a difficult task to load trained dataset in main memory. For that purpose we have removed the redundant datasets. We have trained our dataset from previous data, we did like this because our model should be



trained on previous data and should be able detect fraudulent transaction of the current month, which will help in real world.

## **2.4 CREDIT CARD FRAUDULENT DETECTION USING HIDDEN MARKOV MODEL**

In our paper we utilized HMM to identify fraudulent. We demonstrated the exchanges of MasterCard by utilizing HMM. For swiping reason, we have utilized the RFID gadget to demonstrate the shopping exchanges. We identified the misbehaviors by observing the conduct of the client. We include High security addresses page additionally, in case card is stolen, we have given another profile ID to the consumer and gave ONE TIME PASSWORD for security reasons. We have given right to the admin to block the card from obstructing in case card is lost. As our aim is to achieve the better accuracy but our dataset we could achieve up to 99.97%. As for fraudulent detection, the false alarm plays an important role, as whenever there is a fraud transaction it shows an outlier transaction which will differ from its neighbor or we can say that deviate from the given data point. We give more priority to fraudulent catching algorithm then the false alarm because our aim is to catch the fraudulent at the very first moment.

## **2.5 CREDIT CARD FRADULENT DETECTIONUSING DECISION TREE INDUCTION ALGORITHM**

In Snehal Patiletal, describes the “Decision Tree Induction Algorithm” which is used for Credit Card Fraud Detection <sup>[1]</sup>. In this paper it discusses about the

method, decision tree approach is a new cost sensitive technique compared with well-known traditional classification models on a real-world credit card fraud data set, which reduces the sum of misclassification cost, in selecting the splitting attribute at each of the non-terminal node become advance. Credit card fraud detection is to reduce the bank risks, also used to equalize the transaction information with credit card fraud transaction of historical profile pattern to predict the probability of being fraud on a new online transaction. In this model use of “Credit Card Fraud Detection Using Decision Tree for tracing Email and IP Address. By using this technique, we can able to find out the fraudulent customer/merchant through tracing the fake mail and IP address. If the mail is fake, the customer/merchant is suspicious and information about the owner/sender is traced through IP address.

As prediction of score is much important task according to our model therefore we are predicting the score on the basis of the given formula:

$$\text{Score} = 0.5 * \text{TP} + 0.5 * \text{Deviation}$$

On the basis of these score we made two classes 0 and 1. If the score is 1 it will move to class 1 and termed as legal transaction and if the score is 0 it will move to class 0 and termed as fraudulent transaction. At last, the accuracy is calculated on the basis of how many fraud transactions are there in our dataset and how many we predicted with the help of our model.

## **CHAPTER 3**

### **CREDIT CARD FRAUDULENT DETECTION SYSTEM**

#### **3.1 GENERAL**

Unusual pattern which is known as outliers which not fulfill the expected behaviors is known as Anomaly detection. Many business applications are based upon this technique, unusual patterns in network are identified. Its helps in detecting credit card fraudulent as well as operating system fraudulent. Jupiter notebook we are going to take the credit card fraud detection as the case study so that we can understand the concept in detail. Outlier value is those value which shows an abnormal behavior from its neighbor or we can say that from standard data point. Generally, Outlier data termed as fraudulent transaction. Our experiment based upon catching fraudulent activity with the help of false alarm. Our model has focused on the use of Isolated Random Forest and Local Outlier Factor, however previous works has also been done using Bayesian Regularization and Gradient Descent Adaptive learning algorithms. There are many advantages of this system and one of the major advantage that we are recognizing the pattern and on the basis of pattern we made chart which will help to understand the fraudulent easily because it is easy to understand the data in the form of chart. We have plotted the chart for every features from V1 to V28. We should also keep the things in mind that other financial bank cannot read the other personal information. We can use this technique to find the scheme which

will help in finding credit card fraudulent transaction. The advantages of this system is that it can work in an efficient way for the limited amount of data.

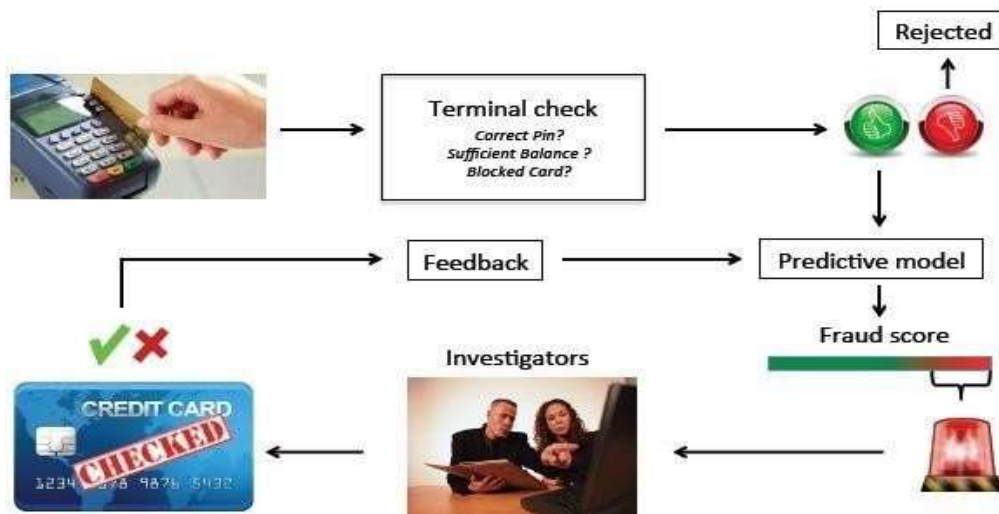
### **3.2 PROBLEM DEFINITION OF CREDIT CARD FRAUDULENT**

As in increase of online payment, increase in use of credit card. Many company provides the facility of credit card payment. We can purchase a lot of things using our credit card. People started doing fraud in this field by using someone's credit card and using someone's personal information to issue credit card. Electronic data can be interchange in case of online payment to perform fraudulent. We cannot prevent credit card fraudulent with the help of credit card billing but we need to prevent the fraudulent also. If we will talk about the success story of all the existing system, it is not that much efficient in finding the fraudulent. So, it's become essential to make a system which can find the fraudulent at the very first time and help customer to reduce the fraudulent in their all online transaction and they can get the notification at the very first time that their credentials are using by some other people. This will help him to overcome with this kind of fraud activity at early and can think to modify their losses. There should be limitation on the credit card that we cannot make transaction above this much amount in on day or at a time. This will reduce the amount of losses.

We have two analyzer as random forest algorithm and local factor outlier which will determine the nature of fraudulent whether it is a legal or fraudulent transaction. These will also help us in calculating the score prediction which will represent a more balancing result.

As the result is the mean between the legal and fraud transaction and then the accuracy is determined on basis of them. The dataset should be more balance because it plays an important role. It's also become a mandatory to split our dataset in train dataset and test dataset.

### 3.3 BLOCK DIAGRAM



### 3.4 METHODOLOGY

The task which is performed for the prediction of transaction and labelled as fraud is detected on the basis of binary classification. We make two class for the prediction of fraud: class 0 and class 1.

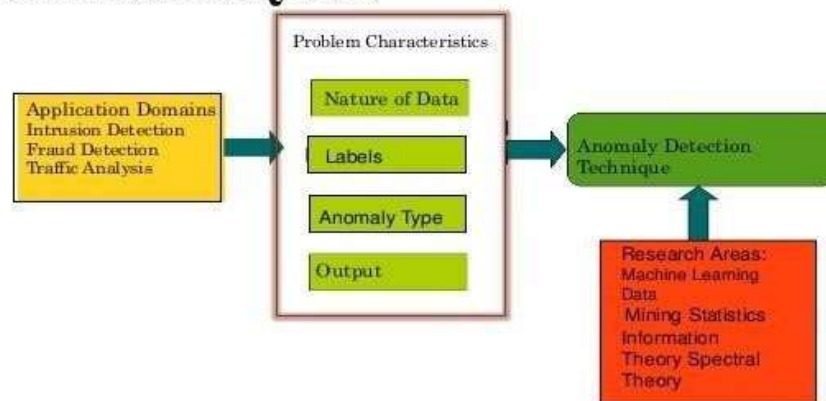
Class 0 if there is no fraud and class 1 to catch the fraud. This can be done with the help of binary classification.

### 3.4.1 WHAT ARE ANOMALIES?

Anomalies can be categorized as following:

- Point Anomalies: Point anomaly is a single instance of data. The credit card fraudulent detection technique is based on “amount spend”.
- Contextual Anomalies: The best example of contextual anomaly is time-series data.
- Collective Anomalies: Here, Detection of anomaly is based on a set of data instances collectively. Therefore, a set of data will help in detecting fraudulent anomaly. If someone try to theft personal data from server it will come under collective anomaly and named as cyber-attack.

## ANOMALY DETECTION TECHNIQUE



## **Figure 3.2: Anomaly Detection Technique**

### **3.4.2 ANOMALY DETECTION**

Identifying an unobserved pattern in new observation is the main area of concern. It's include training of dataset.

### **3.4.3 NOISE REMOVAL**

Noise removal is the process of removing noise from meaningful data, noise is unnecessary data along with the meaningful data.

## **3.5 ANOMALY DETECTION TECHNIQUES**

The various Anomaly Detection Techniques are as follows

### **3.5.1 SIMPLE STATISTICAL METHODS**

The simple way by which we can determine the irregularities in dataset by determining the deviation of data point from common statistical distribution, for example mean, mode and median.

Anomaly data point is that deviates by a certain standard deviation from mean. To compute average data point we need a rolling window across data points which is known as moving average which is used to find low pass filter.

### **3.5.2 CHALLENGES WITH SIMPLE STATISTICAL METHODS**

The low pass filter allows us to identify anomalies in simple use cases, but there are some framework where this method fails to determine anomaly data point. Data which contain noise data which can be named as abnormal data, as the boundary between normal and abnormal are not accurate. Therefore, it's a big problem to identify threshold value because the moving average can't apply in that framework.

### **3.6 CREDIT CARD FRAUDULENT DETECTION SYSTEMS**

All the credit card fraudulent detecting models are evaluated and compared using this model.

**Accuracy** - It is characterized as a bit of all the quantity of exchanges which are distinguished effectively.

**Methodology** - This indicates the instrument pursued by the credit card FDS.

**True Positive or TP** - Legal and fraud transaction are detected on this basis. Genuine transaction only counted here.

**False Positive or FP** - Legal and fraud transaction are detected on this basis. Fraud transaction only counted here.

**Supervised Learning** - In this supervised data is fed in the machine.

### **3.7 FUNCTIONALITIES**

Many organization and banks will take the benefit from this model. Because this will be a significant model for the prediction of credit card fraudulent.



This will detect the consumer behaviors and his last transaction and predict whether the consumer is fraud or not. We use random forest and local outlier factor for the fraudulent. We need to have controls over the algorithm in order to fit with the data set. It will help our application to improve and to be more efficient in order to detect the fraudulent transactions and help us in solving problems.

### **3.8 ACCURACY**

The Fraudulent Detection is done on basis of previous transaction history of consumer. We will detect out of whole transaction how much result in fraud. Then we will identify whether a new transaction made by customer is fraudulent or not. With the help of this model we achieve 99.97% accuracy in finding fraudulent transactions.

### **3.9 OBSERVATION**

The data set contains 492 frauds out of almost 300,000. This results a probability of 17.2% fraudulent cases. This identified that there is much more fraud customer. The data sets consists of column which start from v1 and end as V28. There are much features present from V1 to V28. Furthermore, there is no missing value present in datasets. The datasets has column name as Time & Amount. The analysis is done on the basis of ranges present in this two columns.

The datasets contains the numerical value which can be called as PCA transformation. Due to security issue, unfortunately we cannot take the original features and information about data. Column V1 to V28 are taken as

principal components. The features which is not transformed with PCA are “Time” and “Amount”.

“Time” plays an important role here as it is used to determine the time between each transaction and it is calculated in seconds.

“Amount” is another feature which is used to determine the transactional Amount.

“Class” is the most important feature here in our model which is response variable and it takes the value as 1 and 0. It gives value 1 in case of fraud and value 0 in case of legal transaction. The main goal of this model is to predict the credit card fraudulent, for all transaction which is received as online payment to check whether the transaction is legal or not. If the transaction is genuine then it is consider as legal transaction and the transaction which has fraudulent should be recognize as fraud transaction. All this is performed with the help of random forest algorithm and local outlier factor to make an assumption of true probability and false probability. The result obtain after this algorithm performed successfully is then plotted as graph and heat-map. This model is also tested for different test cases and also compared with the previous all model and the accuracy is also compared.

### **3.10 MODEL PREDICTION**

Now it is time to start building the model. The types of algorithms we are going to use to perform anomaly detection on this data sets are as follows:

### **3.10.1 ISOLATION FOREST ALGORITHM**

Anomalies are detected with the help of Isolation Random Forest algorithm. This algorithm tells the fact that anomalies are data points that are distinct and few. These properties in results describes that, isolation mechanism suspects anomalies. On the basis of above all we came to know that this method is different from all methods which exists in past and more accurate as well. This introduces isolation algorithm is more efficient technique for anomalies detection rather previous algorithm. Moreover, this algorithm takes very less memory and time complexity is also very less. We make binary tree which is small as compare to the datasets.

When both good and bad behaviors present in datasets then Machine learning algorithms should work better to balance the system, and predict the pattern.

### **3.10.2 WORKING PRINCIPLES OF ISOLATED RANDOM FOREST**

The Isolation Random Forest algorithm works by randomly selecting a feature from datasets and then randomly find a split value from minimum and maximum value. According to logic applied, the difference between anomaly observations and normal observation is of few cases. We require more condition in isolating normal observations. The conditions required to differentiate between normal and anomaly observation is used to calculate score.

### **3.10.3 LOCAL OUTLIER FACTOR (LOF)**

Local Outlier Factor (LOF) is an outlier algorithm which provide mechanism to compute the deviation of given data point from its neighbors. It consists

outlier samples which has a low density as compare to its neighbors. The outlier value is chosen on basis of greater and minimum value present in the cluster of datasets and different from its neighbors. If the outlier value is mismatching from its neighbors, then it would have been caught by the system and result in fraudulent It also helps us in finding the deviation of outlier data from the standard deviation which is followed by all the neighbors.

#### **3.10.4 OBSERVATIONS**

- Isolation Forest can detect 73 errors where as Local Outlier Factor can detect 97 errors in order SVM can detect 8516 errors
- Isolation Forest has a better accuracy which is 99.74% than LOF which is 99.65% and SVM has 70.09
- When we compare error precision & recall for these 3 models, the Isolation Forest performance is much better than that of LOF as we can see that the detection of fraud cases is around 27 % in case of Isolation Forest where as in case LOF detection rate of just 2 % and in case of SVM of 0%.

## **CHAPTER 4**

### **INTRODUCTION OF MACHINE LEARNING**

#### **4.1 GENERAL**

AI is a mechanism which features algorithms and calculations based on a normal human intelligence to address a problem. The AI behaves and approaches a problem in a similar way that a normal human brain would. Its working mechanism is influenced by human thinking. A collection of expectation and result is achieved by AI by portraying information in a form termed as 'test information' without making use of any predetermined models or being trained in that particular domain. Problems catering to non-related dimensions such as email sifting, PC vision, location of system gate crashers are addressed. Thus it is assertive that it is not possible to train an AI to address a particular domain, instead an AI trained with general problem solving abilities, builds up its own algorithms for a set of problems.

An AI engine is allocated with responsibility of prediction or analysis using a PC framework and set of data. For this an AI engine is allocated with packages of scientific methods, logistic calculations, data sets and knowledge about the field of the problems for performing.. Moreover, the entire operation of AI is carried based on unsupervised learning model which leaves a very less room for training a robust AI for only a problem specific solution. However, for business purposes modifications are performed before its application.

## 4.2 OVERVIEW OF MACHINE LEARNING

The name was authored in 1959 by Arthur Samuel Tom M. Mitchell gave a generally cited, increasingly formal meaning of the calculations contemplated in the AI field. This meaning of the assignments in which AI is concerned offers an in a general sense operational definition as opposed to characterizing the field in psychological terms. This pursues Alan Turing's proposition in his paper "Registering Machinery and Intelligence", in which the inquiry "Can machines believe?" is supplanted with the inquiry "Can machines do what we (as speculation elements) can do?" In Turing's proposition the different attributes that could be controlled by a reasoning machine and the different ramifications in building one are uncovered.

Before the introduction of machine learning a general assumption was that a robot needs to learn everything from a human brain to function appropriately. But as efforts were made to do so, it was realized that it is very difficult to make a robot to learn everything from a human brain as the human brain is very much sophisticated. An idea was then proposed that rather than teaching a robot everything we know, it is easier to make the robot learn on its own. The type of dataset we are working upon largely determines how we approach while training the model. Based on the dataset we will feed to the algorithm, the training model would vary. The size, type and dynamism of the dataset will decide what type of training model we would build. Finally on deciding upon the training model, modifications need to be made to achieve the proper objective function to generate proper set of output that we wish to achieve. The stages of machine learning process are rather termed as ingredients than

steps, because the machine learning is an iterative process. The iterative process is repeated each time to achieve maximum optimization and efficiency.

### **4.3 MACHINE LEARNING-BASED APPROCHES**

The following is a concise outline of mainstream AI based systems for inconsistency identification.

#### **4.3.1 DENSITY BASED DETECTION OF ANOMALY**

It derives its working mechanism from KNN algorithm

Assumption - Relevant data locates themselves around a common point in close proximity whereas irregular data are placed at a distance. The data points are clustered at a closed proximity based on a density score, which may be derived using Euclidian distance or appropriate methods based on the data. Classification is made on two basis:

K closest neighbor: In this method the basic clustering mechanism is dependent on separation measurements of each data points which determines the clustering or similarities of each information considered.

Relative thickness of the information - Also known as Least Outlier Fraction (LOF).

Calculation is performed on the basis of separation metric.

### **4.3.2 CLUSTERING BASED DETECTION OF ANOMALY**

Clustering is an exceptional algorithm known for its optimization and robust nature. For this reason, it is widely used in unsupervised learning

Assumption - Data points that are similar tends to get gather around specific points. The relative distance of each cluster is achieved by its shortest distance from the centroid of the space.

K means is widely used in data classification. It makes use of k means algorithm to cluster closely related data in close proximity forming clusters.

### **4.3.3 SVM BASED DETECTION OF ANOMALY**

- A support vector machine is one of the most important algorithm used for classification purposes
- The SVM uses methods to determine a soft boundary to distinguish data clusters. Data closely related falls within the parameter of a closed boundary. This results in formation of multiple clusters. SVM is widely used for binary classifications also. Most of the SVM algorithms works based on unsupervised learning.
- The yield of an abnormality locator are mostly numeric scalar qualities for distinguishing areas of explicit edges.

In this Jupiter journal we are going to assume the acknowledgment card misrepresentation recognition as the contextual investigation for



understanding this idea in detail utilizing the accompanying Anomaly Detection Techniques in particular

#### **4.4 DATASET**

A dataset corresponds to a collection of data which may or may not be related to each other. A dataset can consist of data related to a particular domain. It may consist information for a single member or a group of member. For example, personal and other relevant details of an employee can be termed as a dataset, whereas collection of the information of all the employees working for that company is also a dataset. Thus the purpose of the problem defines the size of the dataset. A dataset consists of multiple columns often termed as parameters and multiple rows known as tuples. Individual data pieces are also termed as datum. For example, in a data set consisting of employee details of a company

Datasets which are too large to be operated on by traditional database methods are termed as Big Data. With rising data generation, the need for new algorithms and tools to cope up with thousands of gigabytes of data have given rise to Big Data Analytics. Modified and robust algorithms to optimally operate on the varied range of data is in development. Other than that data is also classified on basis of its dynamisms. A static data requires a single set of algorithms to operate upon whereas a real time data requires a dynamic algorithm to suit the operational needs as and when required.

##### **4.4.1 DATASET DETAILS**

- Time

- Number of seconds slipped by between this exchange and the primary exchange in the dataset
- V1 up to V28
- It might be consequence of a PCA Dimensionality decrease to secure client personalities and touchy features (v1-v28)

#### **4.4.2 AMOUNT**

- Transaction amount
- Class
- 1 for fraudulent transactions, 0 otherwise

# **CHAPTER 5**

## **SYSTEM SPECIFICATION**

### **5.1 GENERAL**

The necessity for the most part dependent on two classes: they are practical portray every single required usefulness for framework administrations which are given by the customers. Non useful necessities characterize the framework properties and compels. The equipment prerequisites indicate the equipment functionalities and required speed and limit of the fringe.

The product prerequisites incorporate programming expected to create and run the framework.

### **5.2 HARDWARE SPECIFICATION**

- System - Core i5
- Mobile - Android
- Monitor - RGB Color
- Hard Disk - 2 TB
- Mouse - Microsoft
- Ram - 8GB

### **5.3 SPECIFICATION OF THE SOFTWARE**

- Operating system - Win 10
- Dataset - csv
- Language - Python

### **5.4 SOFTWARES USED**

- Python 3.5
- NumPy 1.11.3
- Matplotlib 1.5.3
- Pandas 0.19.1
- Seaborn 0.7.1
- SciPy
- Scikit-learn 0.18.1

## **CHAPTER 6**

### **DESIGN ENGINEERING**

#### **6.1 GENERAL**

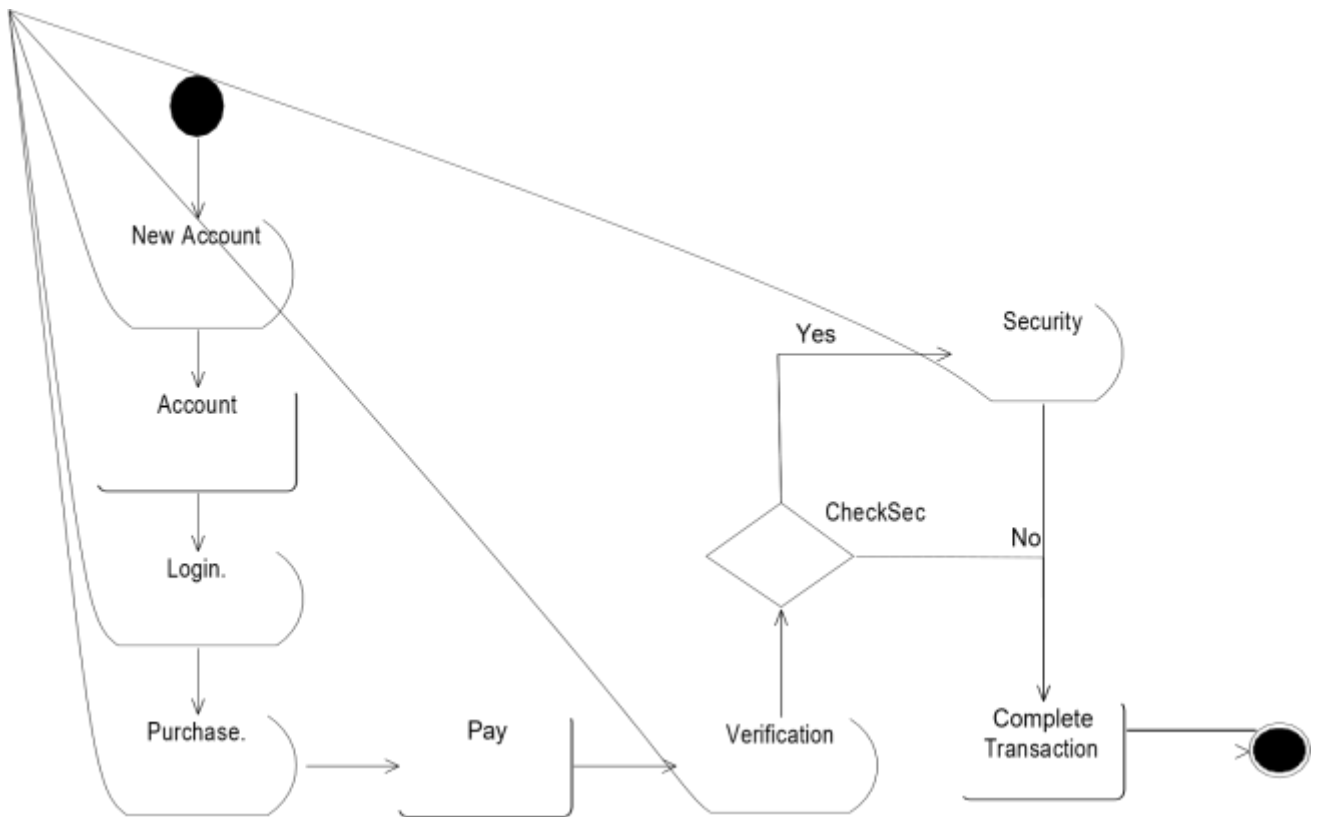
The UML is used for business and production based works. The task of using UML is to provide a solution or working of a product or model using visual representation. UML involves usage of lock diagrams and flow chart to depict the interrelation and workflow of a model. Sometimes it is also used for planning purposes or analysis as a reference for further development of a project

- Provides direction with regards to the requests of the group exercise.
- Software ancient rarities create.
- Directs of errand to individual designers and group.
- Offer the criteria to check & estimate the task's item & exercise.

The UML intestinally process autonomous and can be attached with regards to various procedures. All things considered, It is the most reasonable for utilize driven, intuitive and gradual improvement forms. A case for such procedure is Rational Unified Process (RUP).

## 6.2 ACTIVITY DIAGRAM

It portray the work process conduct of a framework. It ought to be utilized related to other displaying methods, for example, connection and state chart.



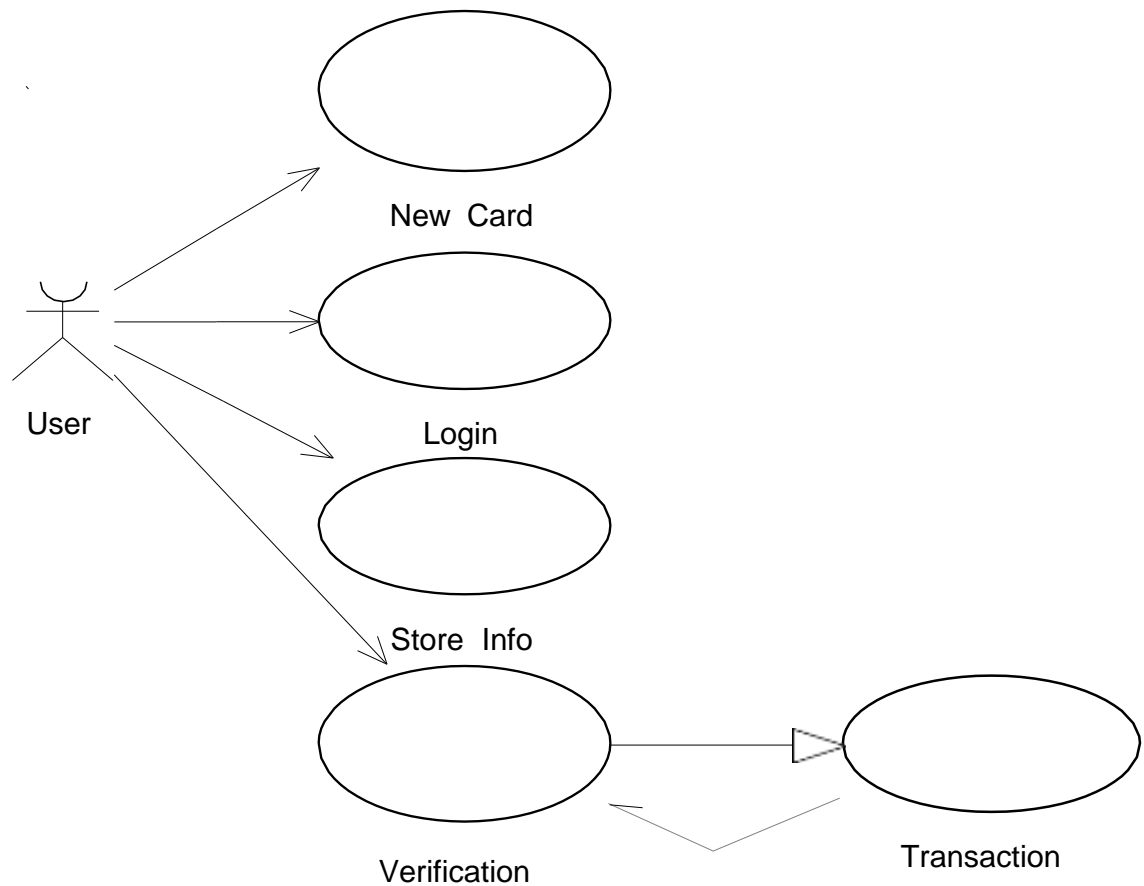
**Figure 6.1: Activity Diagram**

## **6.3 USE CASE DIAGRAM**

Use case chart show the relationship, among performers and clients. Use cases are utilized in pretty much every task they are useful in uncovering prerequisites and arranging the venture. Amid the underlying phase of a task most use cases ought to be characterized yet as a venture proceeds with more become an obvious.

Use case diagrams are used to describe association of actors along with the working model. It is often used to describe a static state of a model. Use case model consists mainly of two components:

Use case charts are<sup>3</sup> formally incorporate into two displaying dialects they are Unified Modelling Language (UML) and System Modelling Language (SML).

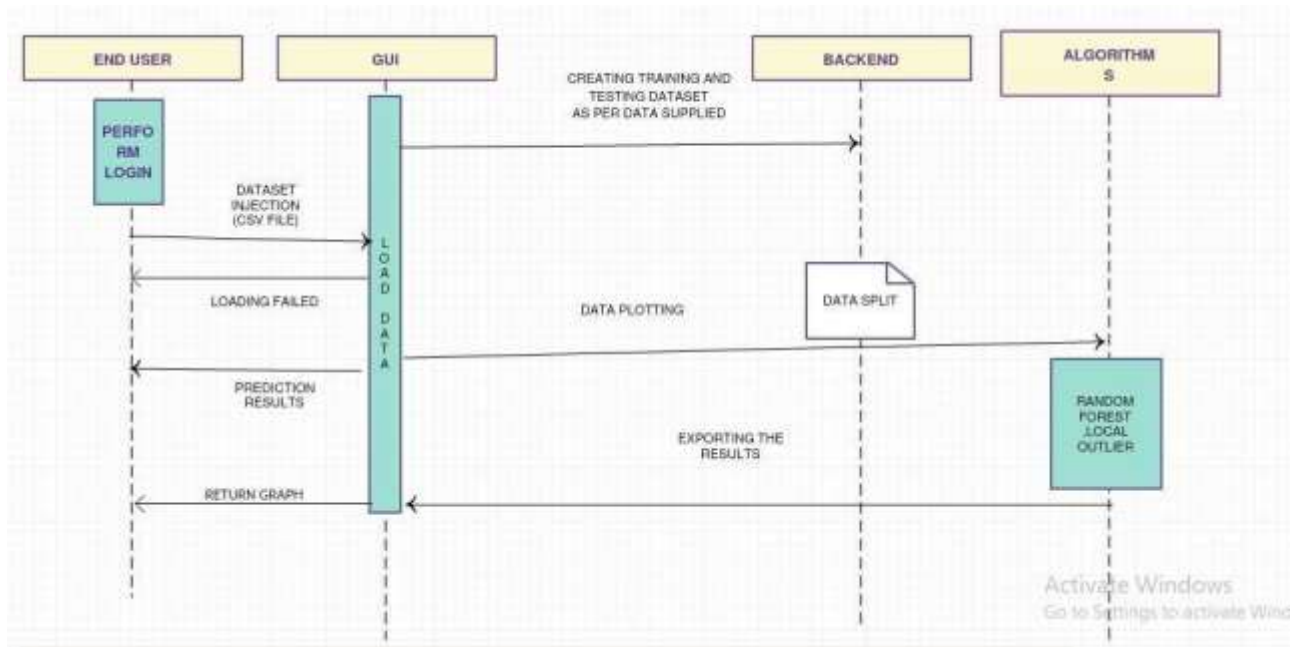


**Figure 6.2: Use Case Diagram**

## **6.4 SEQUENCE DIAGRAM**

Arrangement graph is a collaboration outline that indicates how work with each other and in what request object. Its build of a message arrangement short. A succession outline indicates object connection organized in time arrangement. It delineates the article and classes included the situation and the arrangement of message trade between the items expected to complete the utilitarian of the situation.

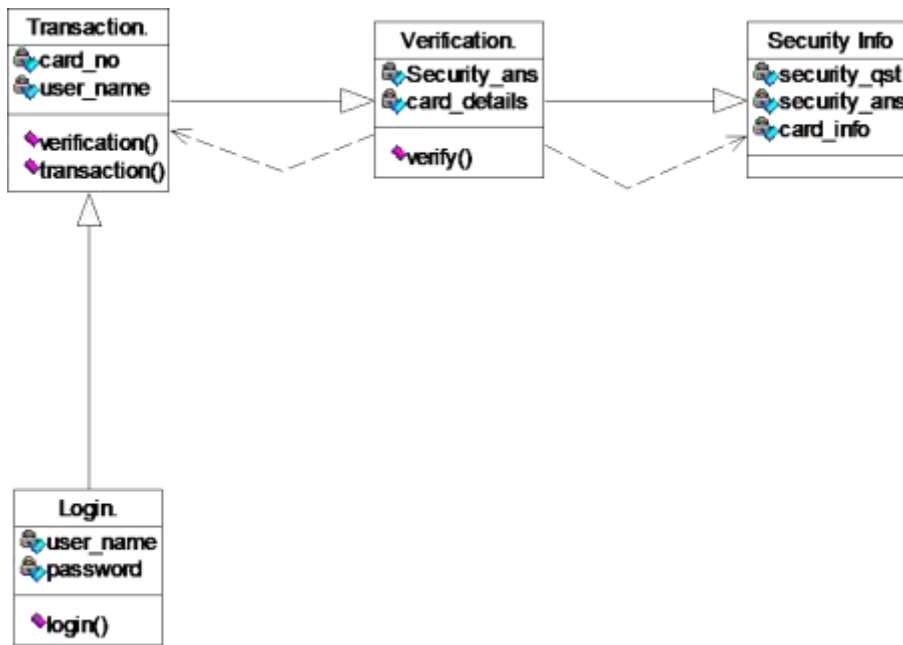




**Figure 6.3: Sequence Diagram**

## 6.5 CLASS FEATURE

Class diagram makes use of inter-related structures which consist of package, entities, objects and variables. This depicts the relationship between each of the entities through associations, containment and inheritance etc. Using a class diagram it becomes easier to understand the holistic working of entities in work along with their inner functionalities. It is widely used in Object Oriented Software designs.

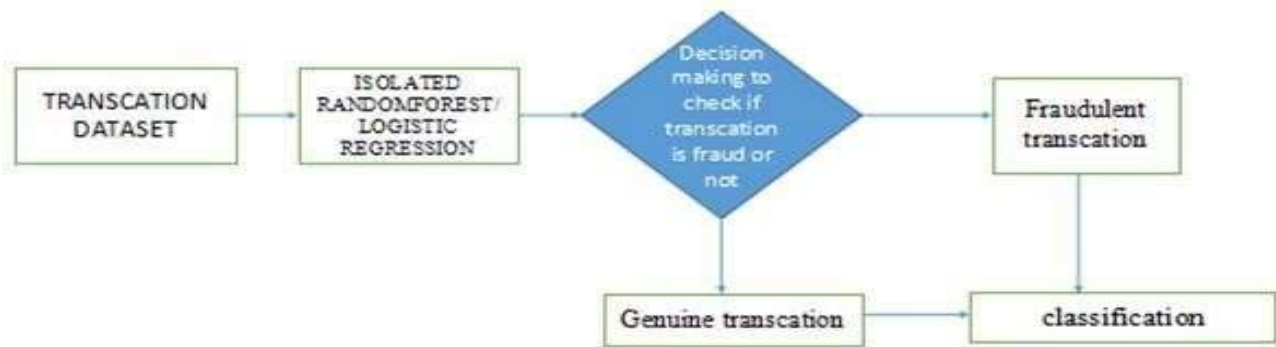


**Figure 6.4: Class Diagram**

## 6.6 THE DATA-FLOW-DIAGRAM

Data Flow Diagram is used to represent the requirements of a system in graphical form. It depicts what the data flow is rather than how they are being processed. It is known as bubbler chart. It defines important transaction in a system as a part of requirement of the model. It is used in the starting phase of a design process for reference of further development on the basis of the current workflow.

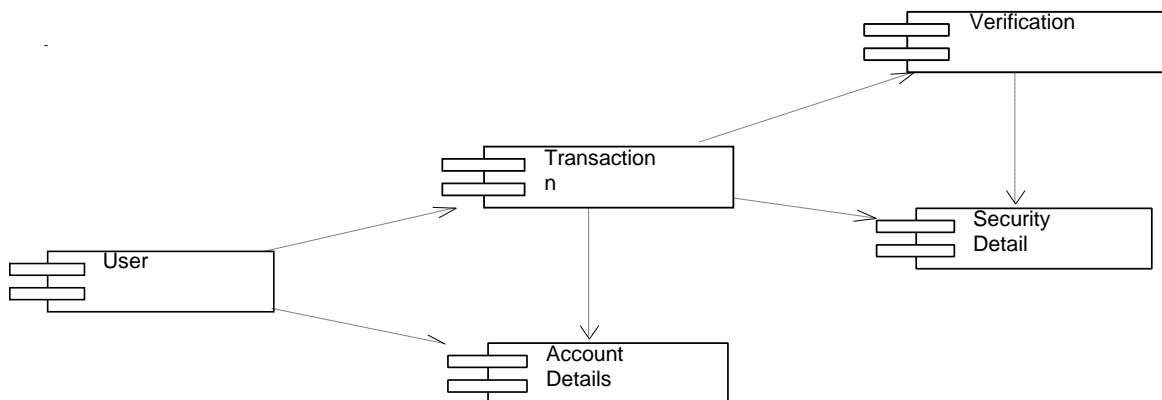
It is depicted by collection of bubbles and lines. The bubbles represents the transactions and operations whereas the lines demonstrates the connection/flow between each transactions. It is independent of hardware, software and datasets used and is a general outline in simple words.



**Figure 6.5: Data Flow Diagram**

## 6.7 COMPONENT DIAGRAM

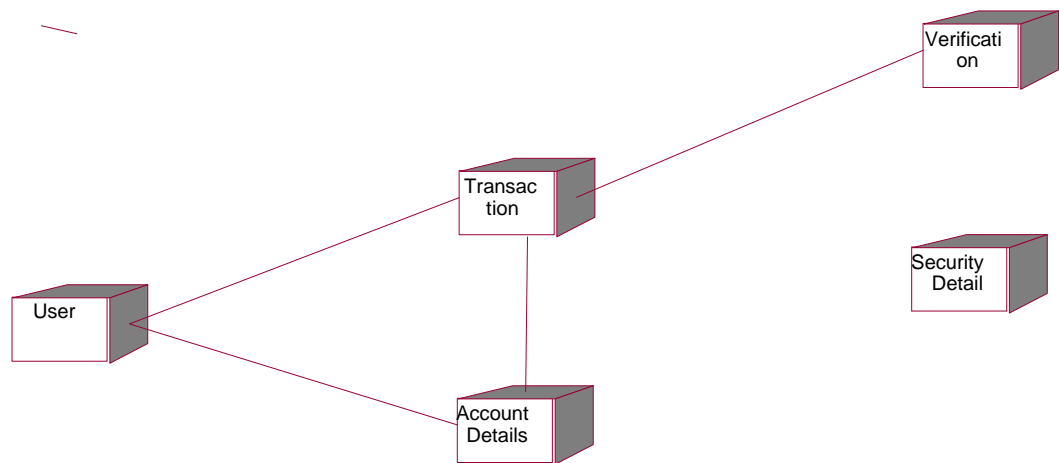
Segment chart shows the abnormal state bundle structure of the code itself. Conditions among parts are demonstrated including source code segments, double code segments, and executable segments. A few parts exist at arrange time, at connection time, at run time well as at more than one time



**Figure 6.6: Component Diagram**

## **6.8 DEPLOYMENT DIAGRAM**

Arrangement graphs shows the design of run time handling components and the product parts, procedures, and items that live on them. Programming part occasion speak to run time appearances of code units.



**Figure 6.7: Deployment Diagram**

# **CHAPTER 7**

## **IMPLEMENTAION**

### **7.1 GENERAL**

Implementation phase brings out the design tweaked out into a operational system. Hence this can be deliberated to be most precarious juncture in accomplishing the efficacious system and in convincing the user faith that system will operate and be effective. This phase encompasses vigilant planning & design, examination of prevailing system and constraints on execution, design & scheming of methods to change over.

### **7.2 PROCEDURE FOLLOWED DURING IMPLEMENTATION**

The application – Credit Card Fraud Detection which is in itself the complete & full-fledged GUI enabled application to envisage/foresee the authenticity & legitimacy of a transaction has been implemented, as per the following steps:

- Install Anaconda from a reliable source.
- Import packages: pandas, Scipy, Matplotlib, Seaborn
- Load the dataset, a dataset is the pool of data for analytical/critical purpose, a (.CSV) file.

- Reconnoiter and get through the dataset through data. Shape, data. Describe.
- Determine the count of fraud cases by checking if class is 0 or 1.
- In the similar procedure, get the correlation matrix.
- Next, there is a need to determine the local outlier factor.
- The GUI is developed using PyQt library.
- The PyQt library, provides tools to achieve a complete GUI enabled application, similar to swings in java environment.
- Define the constructor in the file.

### **7.2.1 DATASET DESIGN**

The dataset holds information about credit card transactions which has been made in a span of two days. The number of frauds have been calculated as 492 out of 284,807 transactions. The details have been given in form of positive and non-positive numerical values. The dataset contains 31 features which has been labelled as V1-V28 due to confidential reasons. The feature which has been revealed are Time and Amount of transaction. Here time denotes the number of seconds elapsed from the first transaction of Day 1. Amount of transaction consists of positive value denoting deposit and non-positive value denoting withdrawal.

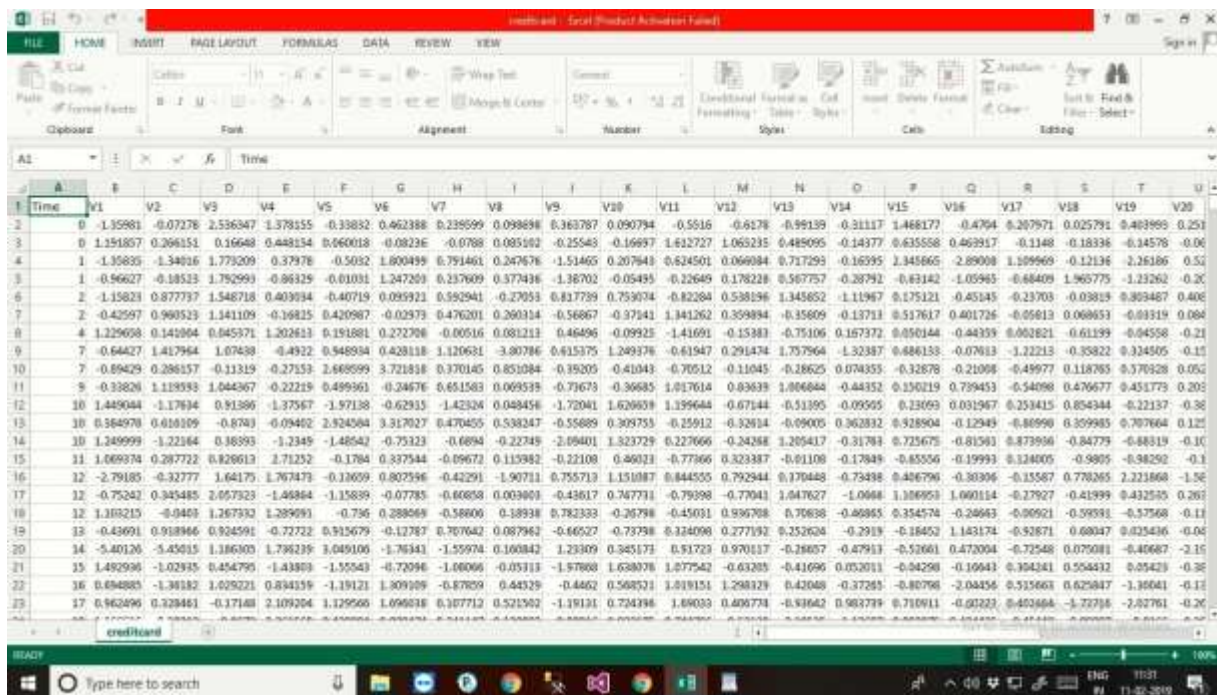


Figure7.1: Dataset Design

## 7.2.2 DATA DESCRIBE

The shape and characteristics of the data values belonging to each column has been described in the following step. The data describe stage belongs to starting stage of exploratory analysis stage.

The figure shows a Jupyter Notebook interface with a code cell. The code cell contains the following Python code:

```
data = pd.read_csv('creditcard.csv')
print(data.shape)
print(data.dtypes)
```

The output of the code is displayed below the code cell. It shows the shape of the data as (10481, 21) and the dtypes for each column. The dtypes are as follows:

Column	Dtype
Time	object
V1	float64
V2	float64
V3	float64
V4	float64
V5	float64
V6	float64
V7	float64
V8	float64
V9	float64
V10	float64
V11	float64
V12	float64
V13	float64
V14	float64
V15	float64
V16	float64
V17	float64
V18	float64
V19	float64
V20	float64

## Figure 7.2: Data Describe

### 7.2.3 PREPROCESSING

The data values has been plotted using histogram describing the numerical distribution of the data values.

After selecting the dataset, the first step is to pre-process the data to make it suitable for model training and testing. In this step, the data were processed in the following ways.

- Finding and filling/removing any null values.
- Standardizing the ‘Amount’ column to make it easy for analysis.
- Removing the ‘Time’ Column from the dataset as it was not contributing much during training and evaluation.
- Checking and removing duplicate entries in the dataset.

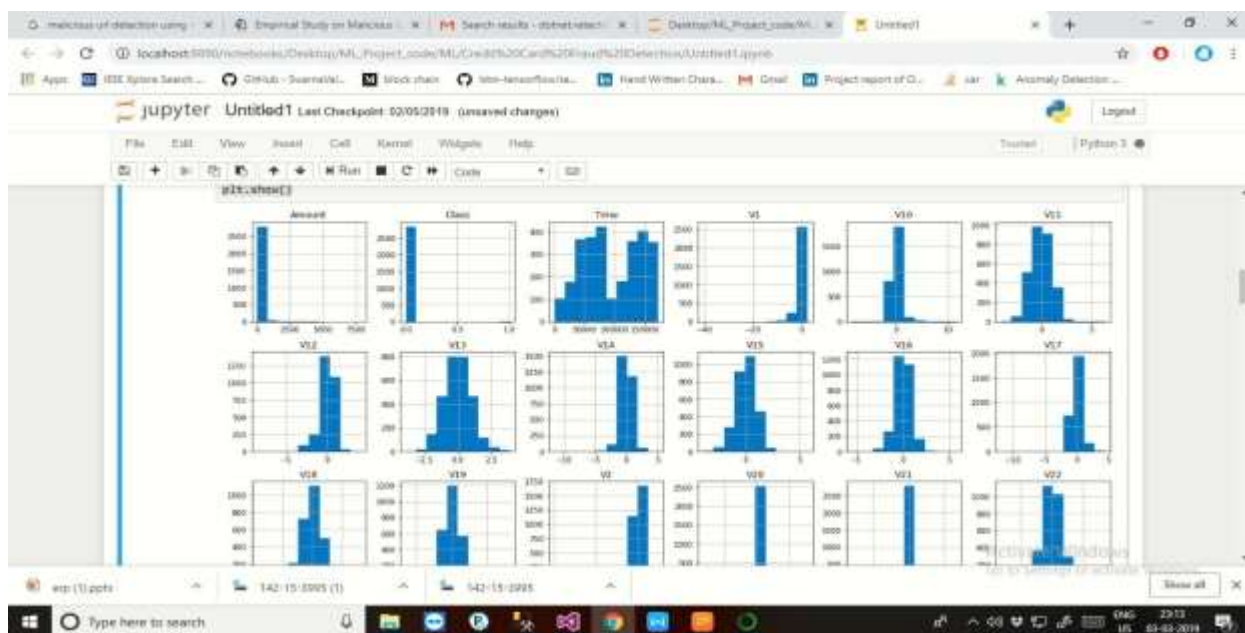
The dataset used in the process was devoid of any missing or null values. It is important to mention that intentional actions to reduce the influence of outliers were not included. The conclusion was based on the understanding that the selected machine-learning model is naturally resistant to outliers. Moreover, incorporating outliers into the dataset was considered advantageous since it brings the model into closer alignment with the complexities of real-world situations. This study sought to improve the model’s capacity to handle the dynamic and different nature of credit card transactions by not using explicit outlier-handling strategies. This approach made the model more adaptable and applicable to real-world scenarios.

The reason for standardising the ‘Amount’ column instead of normalizing it is that, as mentioned in the description of the dataset, all features were the result of



Principal Component Analysis (PCA) except ‘Time’ and ‘Amount’, and the ‘Amount’ scale differed significantly from all other features (V1–V28). Hence, the ‘Amount’ feature was standardized.

Therefore, the use of feature selection techniques was not possible because it would have required clear visibility of feature information. To avoid any potential confusion caused by algorithmic feature selection, a deliberate choice was made to abstain from this process. Furthermore, the ‘Time’ column was excluded from consideration during manual analysis because it did not contribute any meaningful information. It only reflected a sequential count of entries without any temporal significance. Although the lack of feature selection techniques may result in longer training and testing durations, this strategy was considered the best choice to guarantee the retention of all potentially relevant features without relying on feature-specific knowledge.

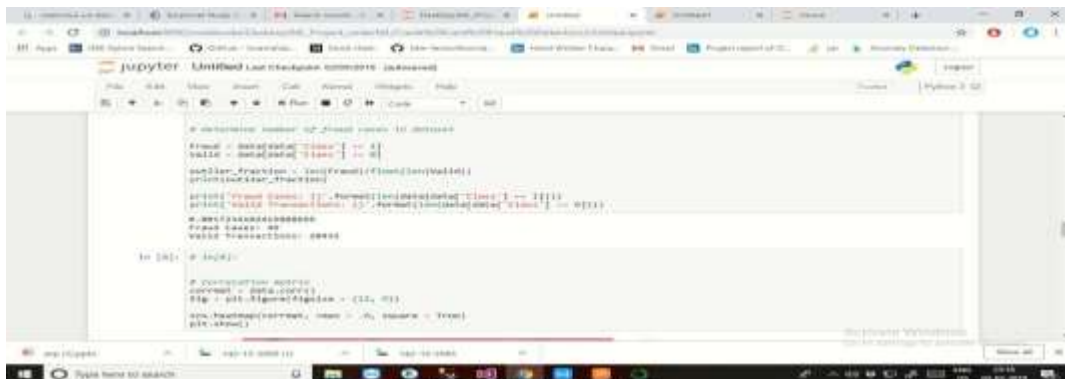


**Figure 7.3: Histogram**

## 7.2.4 FIND FRAUD

Machine learning help with credit card fraud detection? Machine learning models can recognize unusual credit card transactions and fraud. The first and foremost step involves collecting and sorting raw data, which is then used to train the model to predict the probability of fraud.

Detection of the fraudulent transactions will be made by using three machine learning techniques KNN, SVM and Logistic Regression, those models will be used on a credit card transaction dataset.



```
# Importing the dataset
dataset = load_data('data.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values

# Splitting the dataset into the Training set and Test set
from sklearn.cross_validation import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

# Feature Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Fitting the Logistic Regression model to the Training set
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression()
classifier.fit(X_train, y_train)

# Making the predictions on the Test set
y_pred = classifier.predict(X_test)
```

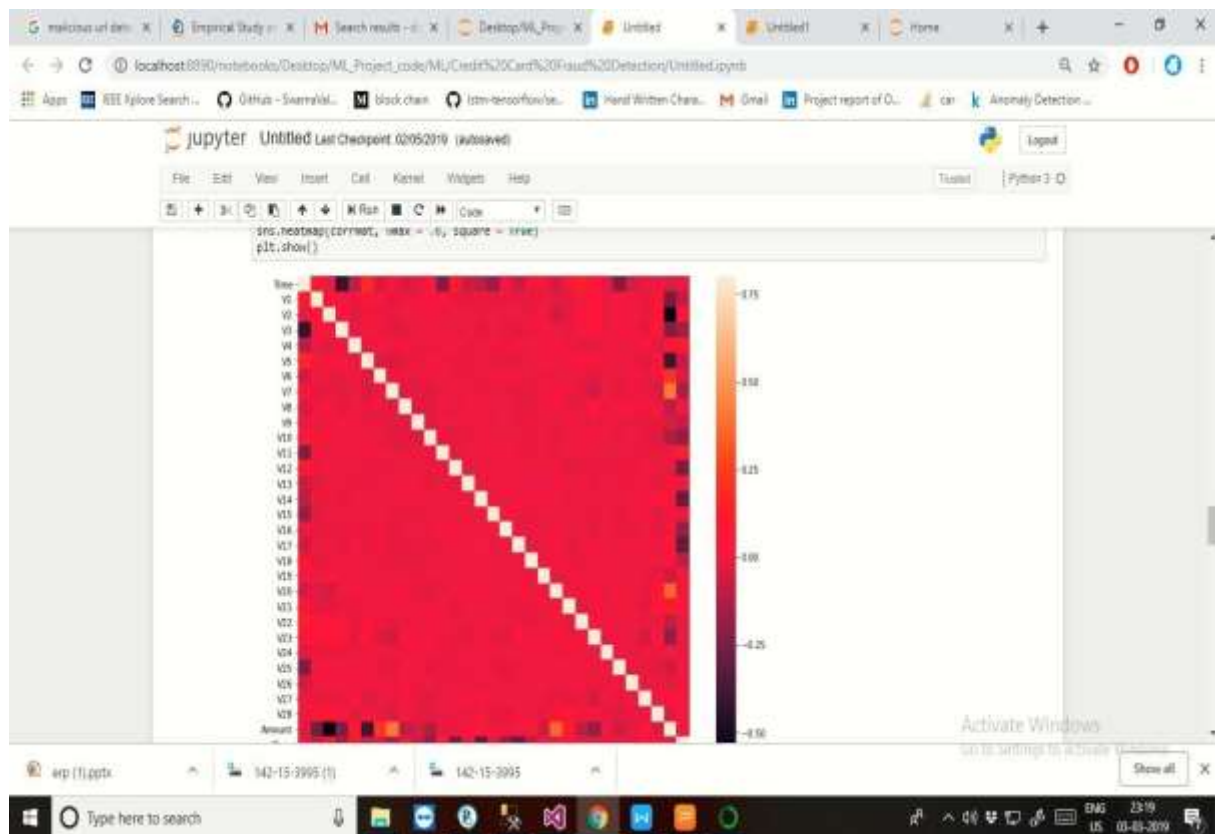
**Figure 7.4:** Fraud Diagram

## 7.2.5 HEATMAP

The heat map has been plotted based on correlation matrix of the features. A correlation matrix describes the relation of features with each other. The level of correlation has been ranged from 0.0-1.0 with 1(white shade) denoting the features to be equilateral and 0(black shade) denoting the features with no interrelation.

A Heat map is a graphical representation of multivariate data that is structured as a matrix of columns and rows.

Heat maps are very useful in describing correlation among several numerical visualizing patterns and anomalies.



**Figure 7.5:** Heat Map Diagram

## 7.2.6 PREDICTION

The prediction that has been achieved using the Isolation Forest Algorithm and Local Outlier Factor Algorithm has been shown below

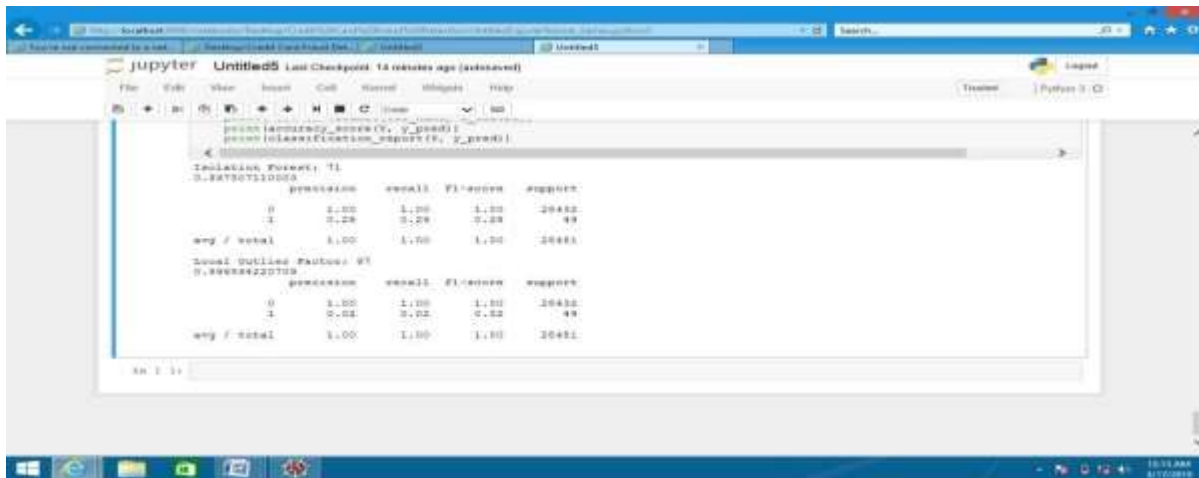


Figure 7.6: Accuracy Diagram

# CHAPTER 8

## SOFTWARE TESTING

### 8.1 GENERAL

In a generalized way, we can say that the system testing is a type of testing in which the main aim is to make sure that system performs efficiently and seamlessly. The process of testing is applied to a program with the main aim to discover an unprecedented error, an error which otherwise could have damaged the future of the software. Test cases which brings up a high possibility of discovering and error is considered successful. This successful test helps to answer the still unknown

### 8.2 TESTING

**Table 8.1:** Tabulated Results

Test Case (sample split)	Assumption	Description	Expected Output	Actual Output		Log Message
				Isolation Forest Algorithm - Algorithm I Accuracy (%)	Local Outlier Factor - Algorithm II Accuracy (%)	
10:90	Algorithm-I will perform better	Check for accuracy at 10%	99.70505	99.75071	99.65942	Success

		training of data				
15:85	Algorithm-II will perform better	Check for accuracy at 15% training of data	99.71675	99.75421	99.67931	Fail
20:80	Algorithm-II will perform better	Check for accuracy at 20% training of data	99.73485	99.69628	99.77352	Success
25:75	Algorithm-I will perform better	Check for accuracy at 25% training of data	99.73311	99.77107	99.69523	Success
30:70	Algorithm-I will perform better	Check for accuracy at 30% training of data	99.73425	99.77645	99.69218	Success

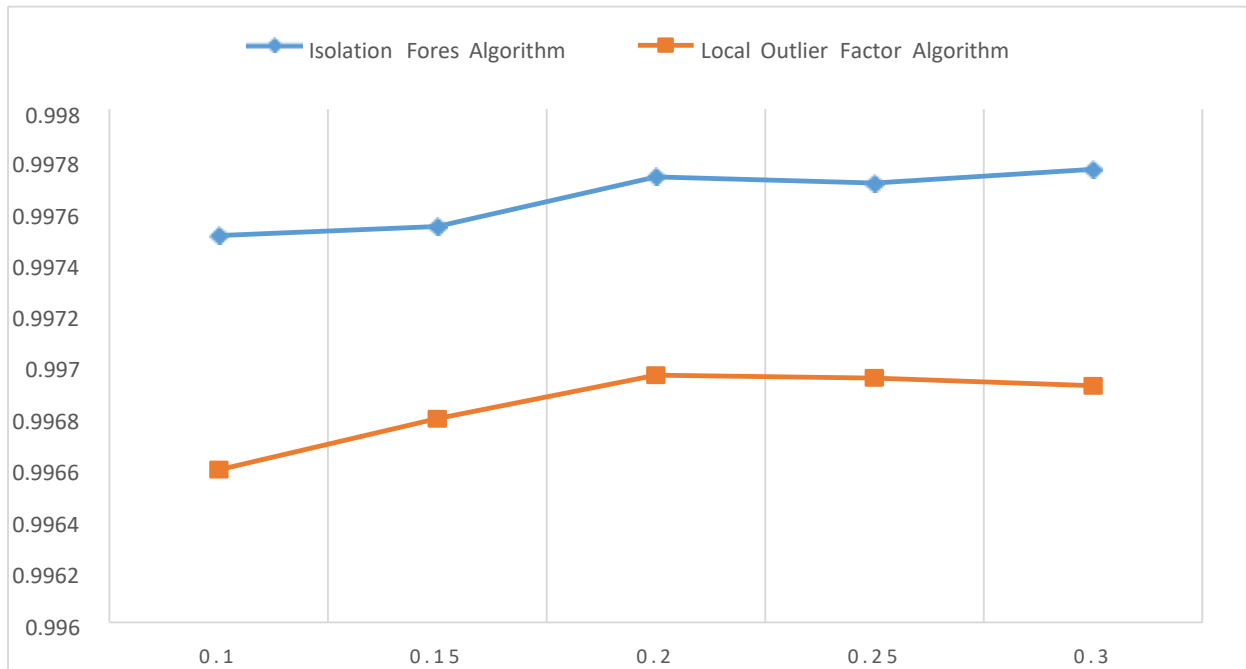
The test cases has been based on the following sample split (train: test) :- (10:90), (15:85), (20:80), (25:75) and (30:70).

**Outlier Fraction:** Describes the ratio of outlier values to the real values in the dataset

**Data Shape:** Describes the number of rows and columns in the training sample.

**Isolation Forest Algorithm Accuracy:** Describes the accuracy achieved on the test dataset using Isolation Forest Algorithm

**Local Outlier Factor Accuracy:** Describes the accuracy achieved on the test dataset using Local Outlier Factor



**Figure 8.1: Comparison Chart**

As we tested the application under different test conditions, the application gave appropriate results. The above chart depicts the accuracy based on two algorithms used, i.e. the Isolation Forest Algorithm and the Local Outlier Factor Algorithm.

## **CHAPTER 9**

### **CONCLUSION**

In this model, we discussed about credit card fraud detection using machine learning. The proposed model has been extensively tested on different types of transactions. The results were promising, almost all the fraudulent transactions could be detected successfully, and the proposed methodology has been compared with existing method and the results shows that proposed method performs superior than existing methods.

In this model, we detected the fraudulent transactions and recognized which illustrates the robustness of the proposed system. This proposed model took the trained dataset and performed classification on basis of them, if the transaction was legal then it moved to class 0 and if the transaction was fraud then it moved to class 1, and significantly improve the detection accuracy.

The proposed method works efficiently in various platform, vivid environment and is a full- fledged cross platform application. The system has depicted robust, scalable and accurate performance to the degree that efficiency is taken into consideration in the Credit Card Fraud Detection System.

The system takes into consideration various factors and has been fulfilling or meeting all the project specifications documented.



# **CHAPTER 10**

## **ALGORITHM**

### **10.1 Logistic Regression**

Logistic Regression is a Classification model, which tries to classify the data based on the probability of it occurring. This algorithm is used in multiple places where classification is required, we have used it to classify if the patient is susceptible to be infected by covid or not this is one of the classification methods which we have used. It used sigmoid function to classify the data. Logistic regression is a supervised machine learning algorithm that accomplishes binary classification tasks by predicting the probability of an outcome, event, or observation. The model delivers a binary or dichotomous outcome limited to two possible outcomes: yes/no, 0/1, or true/false. Logical regression analyzes the relationship between one or more independent variables and classifies data into discrete classes. It is extensively used in predictive modeling, where the model estimates the mathematical probability of whether an instance belongs to a specific category or not. For example, 0 – represents a negative class; 1 – represents a positive class. Logistic regression is commonly used in binary classification problems where the outcome variable reveals either of the two categories (0 and 1).

### **10.1.1 ADVANTAGES OF LOGISTICS ALGORITHM**

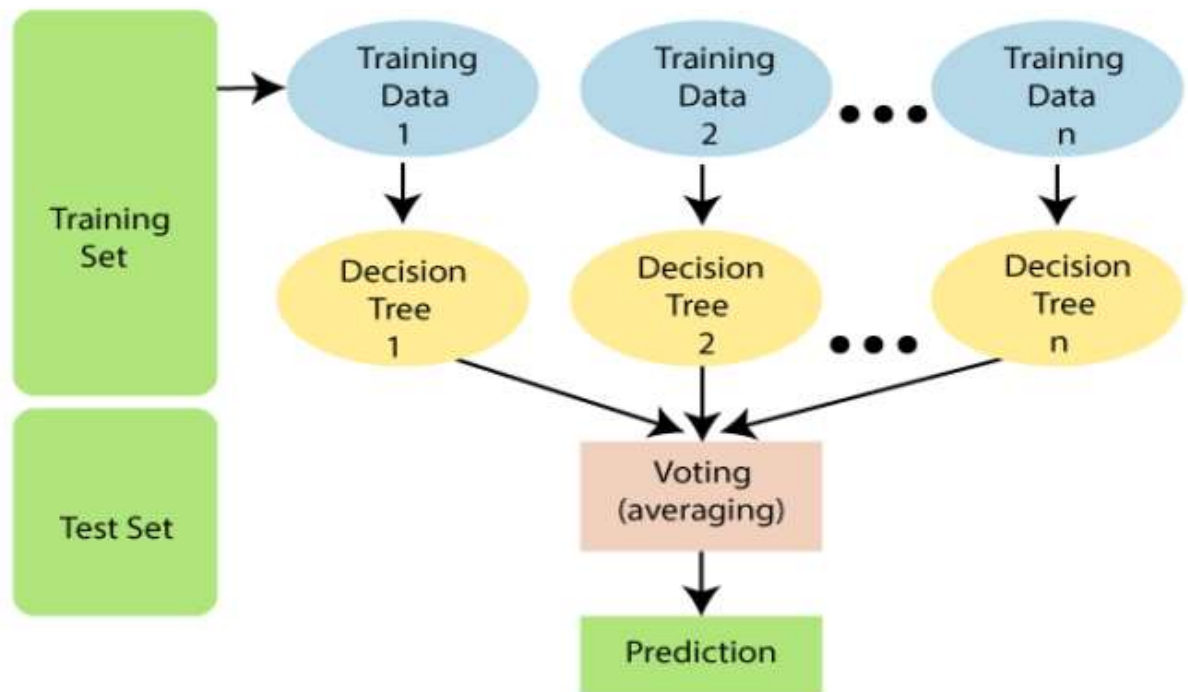
The logistic regression analysis has several advantages in the field of machine learning.

1. Easier to implement machine learning methods: A machine learning model can be effectively set up with the help of training and testing. The training identifies patterns in the input data (image) and associates them with some form of output (label). Training a logistic model with a regression algorithm does not demand higher computational power. As such, logistic regression is easier to implement, interpret, and train than other ML methods.
2. Suitable for linearly separable datasets: A linearly separable dataset refers to a graph where a straight line separates the two data classes. In logistic regression, the y variable takes only two values. Hence, one can effectively classify data into two separate classes if linearly separable data is used.
3. Provides valuable insights: Logistic regression measures how relevant or appropriate an independent/predictor variable is (coefficient size) and also reveals the direction of their relationship or association (positive or negative).

### **10.2 RANDOM FOREST CLASSIFIER**

Random forest is a supervised learning algorithm. The "forest" it builds is a group of decision trees, usually trained with the “bagging” system. The general idea of the bagging system is that a combination of learning models increases the overall result. Put simply: random forest builds multiple decision trees and combines them together to get a more accurate and stable prediction. One big advantage of random forest is that it can be used for both classification and regression problems, which form the most of current

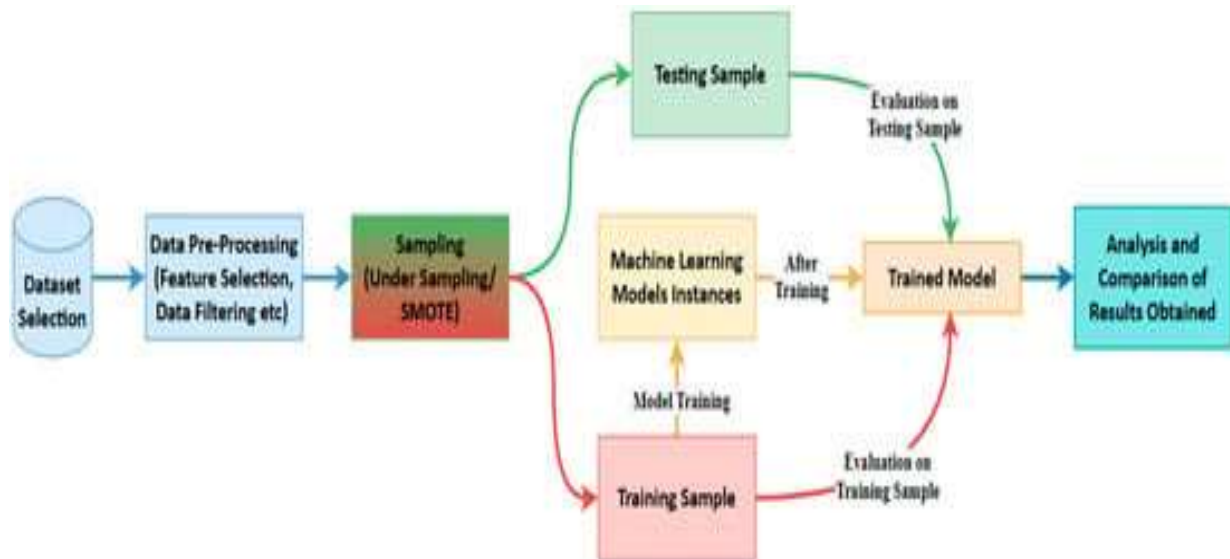
machine learning systems. 39 A Random Forest Algorithm is a supervised machine learning algorithm that is extremely popular and is used for Classification and Regression problems in Machine Learning. We know that a forest comprises numerous trees, and the more trees more it will be robust. Similarly, the greater the number of trees in a Random Forest Algorithm, the higher its accuracy and problem-solving ability. Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. It is based on the concept of ensemble learning which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model.



# CHAPTER 11

## IMPLEMENTATION

Machine learning detects fraud by leveraging historical data on both fraudulent and non-fraudulent transactions. ML algorithms excel at identifying abnormalities in transactions before they escalate into unmanageable issues. Illustrates the flow diagram depicting how machine learning detects credit card fraud.



The architecture of the implementation process, encompassing dataset pre-processing and the division of the dataset into training and testing data. The training dataset is subsequently input into the chosen models for both the training and testing phases. Following this, the evaluation and results are conducted on the trained model to assess its performance.



**Figure 3. The Architecture of the Proposed Credit Card Fraud Detection Model.**

Algorithm 1 presents the algorithm of the proposed model. The algorithm follows a systematic approach, starting with data loading, preprocessing, and sampling. The numbers in explains the sequence of each phase in the whole process. It then iterates over different machine learning model types, trains each model, and evaluates their performance using testing data. Finally, the results, including confusion matrices, are displayed for analysis.

As shown in the initial step in the process involves selecting a dataset containing records of both legitimate and fraudulent transactions. Due to the presence of unordered, raw, missing, or duplicate instances in the dataset, system predictions may be prone to inaccuracies, requiring data pre-processing. To address data imbalance, the sampling of imbalanced datasets is performed. Subsequently, the organized and sampled data are divided into training and testing samples, where the chosen machine learning models are trained using the training sample, and both samples are employed to observe the behavior of the trained models.

## **CHAPTER 12**

### **APPLICATION AND FUTURE ENHANCEMENT**

#### **12.1 GENERAL**

Implementation is the most critical phase to attain a fruitful system and providing the users assurance that the new system is feasible and operative. Each module is tried and tested discretely using the data and substantiated in the mode indicated as per program specification, system and the environment is tested as per user requirement.

The frequently techniques for fraud detection are Nave Bayes, support vector machine and the k - nearest neighbor algorithm. Here, this document has trained various machine learning practices and techniques used in detection of fraud in credit card and assess each methodology based on certain design measures and criteria.

Nevertheless, if there is a need to contrivance a platform that performs real-time credit card fraud detection, it is imperious to reach precisions of 95%, as the odds of false positives along with false negatives is else quite elevated to be used for business application. Impending task must subsequently be focused at exploring further relevant features to enhance, execution of a thorough optimization, and doing real-time tests. Other than the major fraud practices some other types of frauds are done such as through phishing, skimming, credit card generator etc.<sup>[6]</sup>. Also the possibility regrettably not

pursued for timing issues is to refine the metrics in form of commercial forfeiture resolution system, the tenacity of model wouldn't be to capitalize on the count of transactions precisely organized, but instead minimize the costs associated with following up on fraudulent transactions based on the confidence of the model and the associated financial loss. Finally, approaches for dealing with the 'refused' examples are to be explored.

## **12.2 FUTURE ENHANCEMENTS**

There is a very strong possibility of the system being adopted as a norm for the major banking and financial services applications as fraud detection and prevention is the major checkpoint in financial and banking sector.

The above system is also likely to be embedded in other applications based, modified as per platform-specific/application specific environment. The banks, financial and retail institutes have faced huge losses owing to cause of a robust and accurate system to predict and prevent the fraudulent transactions going on in an institution.

This in-turn affects the business capabilities and consumer trust of the company.

Thus, the organizations have moved their focus onto implementing a system which can depict inconsistent transactions, providing banks a privilege to act upon it take necessary measures.

## CHAPTER 13

### SOURCE CODE

```
import pandas
import matplotlib
import seaborn
import scipy
print('Python: {}'.format(sys.version))
print('Numpy: {}'.format(numpy.__version__))
print('Pandas: {}'.format(pandas.__version__))
print('Matplotlib: {}'.format(matplotlib.__version__))
print('Seaborn: {}'.format(seaborn.__version__))
print('Scipy: {}'.format(scipy.__version__))
```

#### **#import the necessary libraries**

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

#### **# Load the dataset from the csv file using pandas**

```
data = pd.read_csv('creditcard.csv')
```



**# Start exploring the dataset**

```
print(data.columns)
```

**#Print the shape of the data**

```
data = data.sample(frac=0.1, random_state = 1)
```

```
print(data.shape) print(data.describe())
```

**# Plot histograms of each parameter**

```
data.hist(figsize = (20, 20)) plt.show()
```

**# Determine number of fraud cases in dataset**

```
Fraud = data[data['Class'] == 1]
```

```
Valid = data[data['Class'] == 0]
```

```
outlier_fraction = len(Fraud)/float(len(Valid))
```

```
print(outlier_fraction)
```

```
print('Fraud Cases: {}'.format(len(data[data['Class'] == 1])))
```

```
print('Valid Transactions: {}'.format(len(data[data['Class'] == 0]))) from  
sklearn.metrics
```

```
import classification_report, accuracy_score from sklearn.ensemble
```

```
import IsolationForest from sklearn.neighbors
```

```
import LocalOutlierFactor
```

**# define random states state = 1**

**# define outlier detection tools to be compared classifiers**

```
for i, (clf_name, clf) in enumerate(classifiers.items):
    if clf_name == "Local Outlier Factor":
        y_pred = clf.fit_predict(X)
        scores_pred = clf.negative_outlier_factor_
    else:
        clf.fit(X)
        scores_pred = clf.decision_function(X)
        y_pred = clf.predict(X)
        # Reshape the prediction values to 0 for valid, 1
        for fraud. y_pred[y_pred == 1] = 0 y_pred[y_pred == -1] = 1 n_errors =
        (y_pred != Y).sum()
# Run classification metrics
print('{}: {}'.format(clf_name, n_errors))
print(accuracy_score(Y, y_pred))
print(classification_report(Y, y_pred))
```

## **GRAPHICAL USER INTERFACE CODE**

```
import sys
import numpy
import pandas
import matplotlib.pyplot as plt
import seaborn
import scipy
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot
import seaborn as sns

self.predictBtn.clicked.connect(self.predictFn)
self.predictBtn.setFont(self.fontLiber1)

self.figure1 = plt.figure()
self.canvas1 = FigureCanvas(self.figure1)
self.toolbar = NavigationToolbar(self.canvas, self)
self.button1 = QPushButton('Plot')    self.button1.clicked.connect(self.plot1)

self.figure2 = plt.figure()
self.canvas2 = FigureCanvas(self.figure2)
self.toolbar = NavigationToolbar(self.canvas, self)
self.button2 = QPushButton('Plot')    self.button2.clicked.connect(self.plot2)

self.layout1=QtWidgets.QGridLayout()
self.layout1.addWidget(self.canvas1,0,0,QtCore.Qt.AlignCenter)
self.layout1.addWidget(self.button1,0,1,QtCore.Qt.AlignCenter)
self.layout1.setHorizontalSpacing(40)
```

```
self.layout2=QtWidgets.QGridLayout()
self.layout2.addWidget(self.canvas2,0,0,QtCore.Qt.AlignCenter)
self.layout2.addWidget(self.button2,0,1,QtCore.Qt.AlignCenter)
self.layout2.setHorizontalSpacing(40)
```

### **# set the layout**

```
self.layout = QtWidgets.QHBoxLayout()
self.layout.addLayout(self.layout1)
self.layout.addLayout(self.layout2)
```

```
self.mainLayout=QtWidgets.QGridLayout()
self.mainLayout.addWidget(self.titleBox,0,0,QtCore.Qt.AlignCenter)
self.mainLayout.addLayout(self.uploadLayout,1,0,QtCore.Qt.AlignCenter)
self.mainLayout.addLayout(self.layout,2,0,QtCore.Qt.AlignCenter)
```

```
self.mainWidget=QtWidgets.QWidget()
self.mainWidget.setLayout(self.mainLayout)
self.setCentralWidget(self.mainWidget)
self.varText=1
```

```

def plot1(self):
# data = [random.random()
for i in range(10)]
self.data.hist(figsize = (20, 20))
# plt.show()
self.figure.clear()

ax = self.figure.add_subplot(111)
ax.plot(self.data, '*-')
self.canvas.draw()

def plot2(self):
pass

def uploadFn(self):
self.fname = QtGui.QFileDialog.getOpenFileName(self, 'Open file','', "Image files
(*.csv *.CSV ")
if(self.fname):
self.data = pd.read_csv(self.fname)
print(self.data.columns)
QtGui.QMessageBox.information(self, "Message", " File Upload successfully ")

```

```

def predictFn(self):
self.data.hist(figsize = (20, 20))
plt.show()
self.data = self.data.sample(frac=0.1, random_state = 1)
print(self.data.shape)
print(self.data.describe())

Fraud = self.data[self.data['Class'] == 1]
Valid = self.data[self.data['Class'] == 0]

outlier_fraction = len(Fraud)/float(len(Valid))
print(outlier_fraction)

print('Fraud Cases: {}'.format(len(self.data[self.data['Class'] == 1])))
print('Valid Transactions: {}'.format(len(self.data[self.data['Class'] == 0])))

corrmat = self.data.corr()
fig = plt.figure(figsize = (12, 9))

sns.heatmap(corrmat, vmax = .8, square = True)
plt.show()

```

```
# Get all the columns from the dataFram
columns = self.data.columns.tolist()
X = self.data[columns]
Y = self.data[target]

# Print shapes
print(X.shape)
print(Y.shape)
```

```
from sklearn.metrics
import classification_report, accuracy_score
from sklearn.ensemble
import IsolationForest
from sklearn.neighbors
import LocalOutlierFactor
```

```
# define outlier detection tools to be compared
```

```
classifiers = {"Isolation Forest": IsolationForest(max_samples=len(X),
contamination=outlier_fraction, random_state=state),
```

```
# Fit the model
```

```

plt.figure(figsize=(9, 7))

n_outliers = len(Fraud)

for i, (clf_name, clf) in enumerate(classifiers.items()):

    # fit the data and tag outliers
    if clf_name == "Local Outlier Factor":
        y_pred = clf.fit_predict(X)
        scores_pred = clf.negative_outlier_factor_
    else:
        clf.fit(X)
        scores_pred = clf.decision_function(X)
        y_pred = clf.predict(X)

    # Reshape the prediction values to 0 for valid, 1 for fraud.
    y_pred[y_pred == 1] = 0
    y_pred[y_pred == -1] = 1

    n_errors = (y_pred != Y).sum()

    # Run classification metrics
    print('{}: {}'.format(clf_name, n_errors))
    print(accuracy_score(Y, y_pred))

```



```
# print(classification_report(Y, y_pred))
```

```
#plt.show()
```

```
if __name__ == '__main__':
```

```
    currentApp = QtWidgets.QApplication(sys.argv)
```

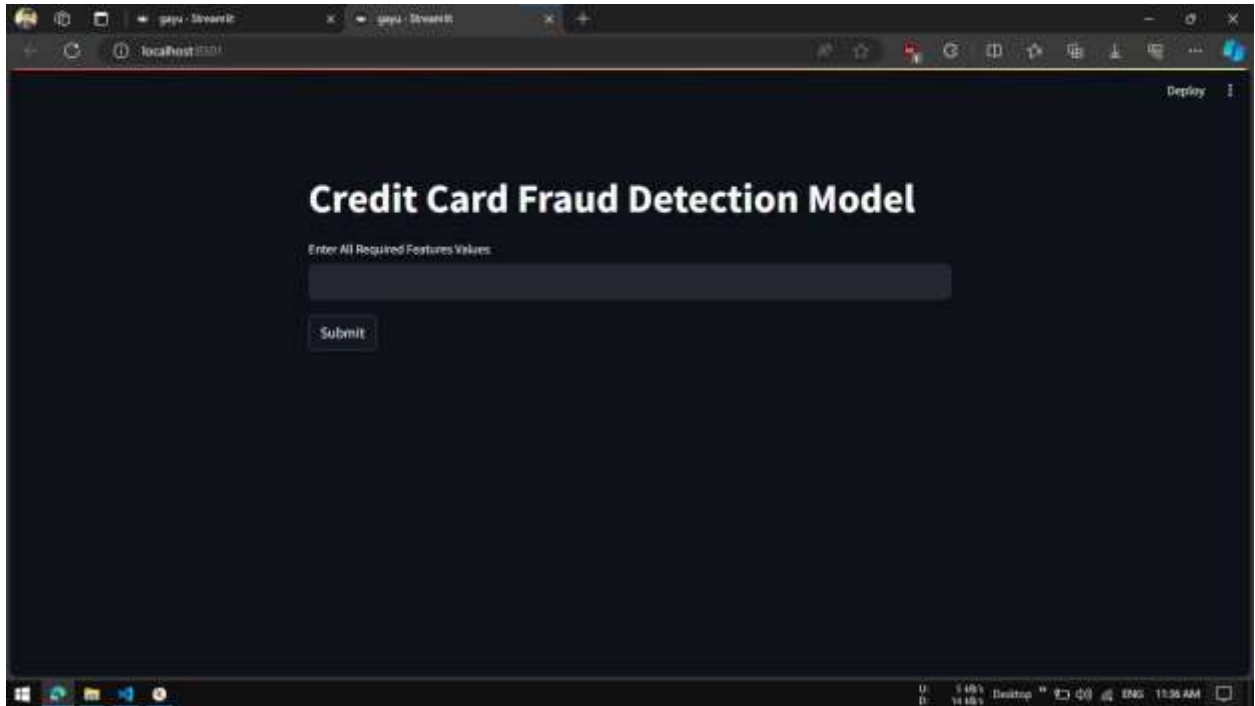
```
    currentWindow.show()
```

```
    sys.exit(currentApp.exec_())
```

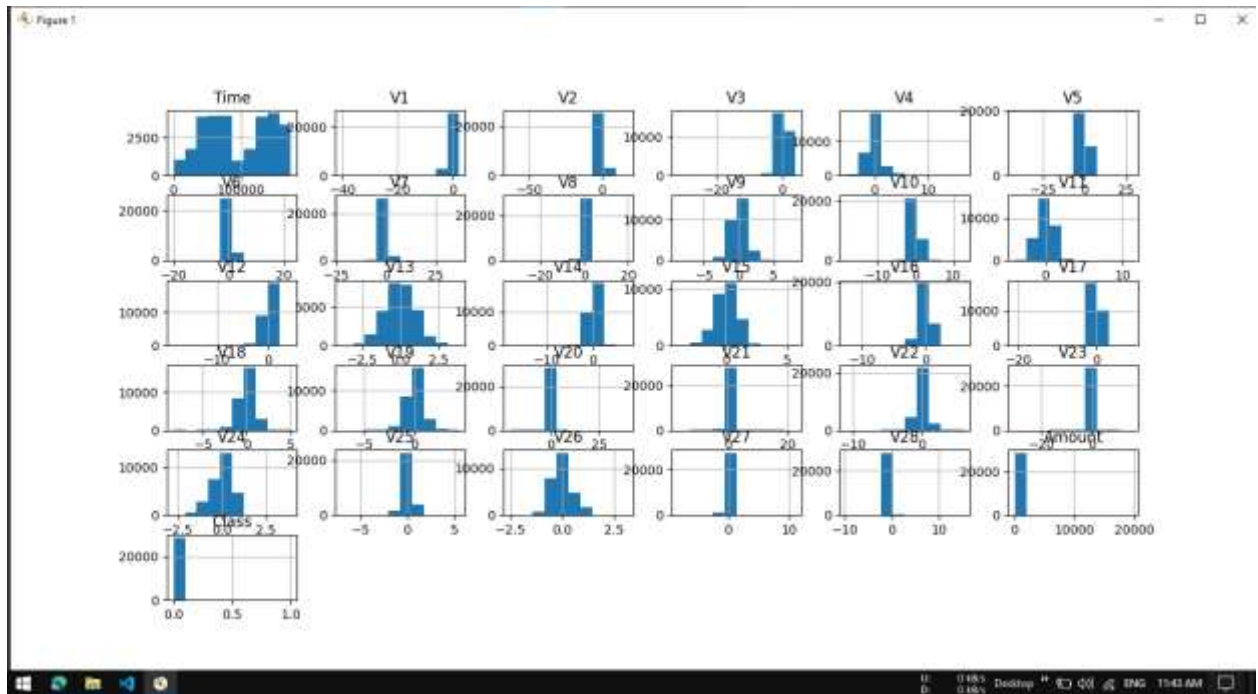
# CHAPTER 14

## OUTPUT

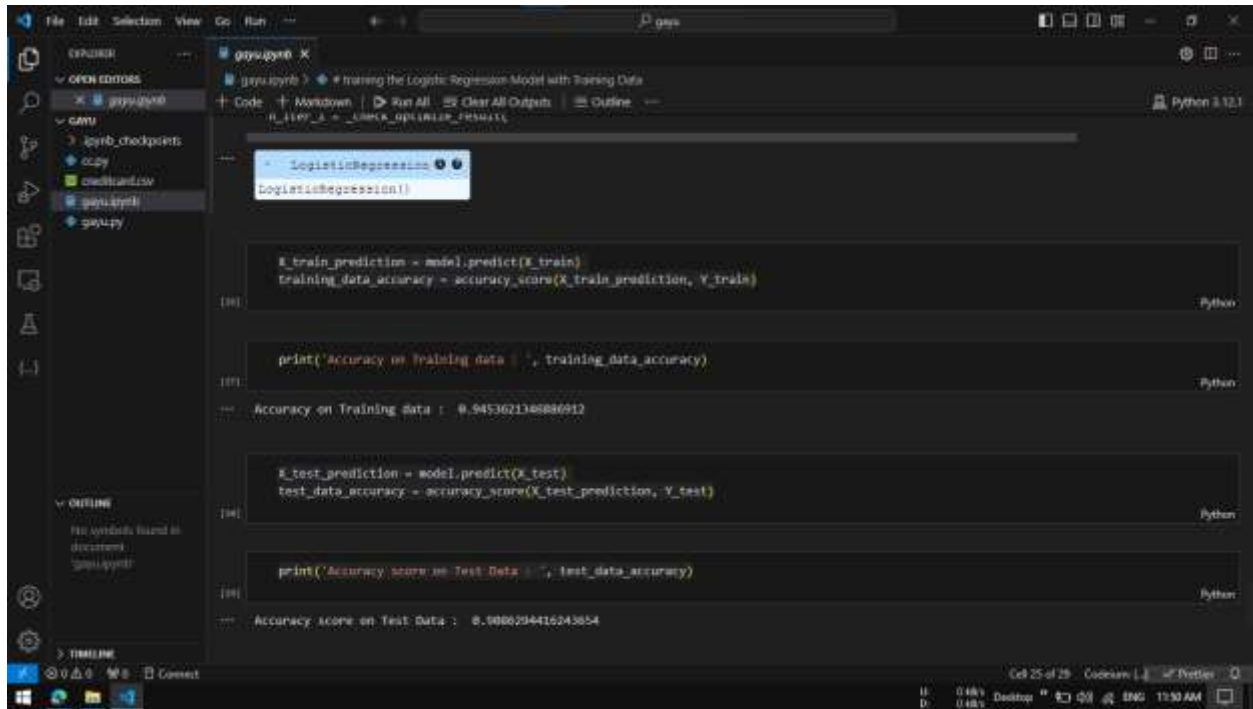
### 14.1 FRONT END FRAMEWORK



## 14.2 DATA PREPROCESSING



## 14.3 ACCURACY



The screenshot shows a Jupyter Notebook interface with a dark theme. The left sidebar contains a file explorer and a list of open notebooks. The main area displays a Python cell with the following code:

```
from sklearn.linear_model import LogisticRegression

# Training the Logistic Regression Model with Training Data
X_train = X_train.astype('float32')
y_train = y_train.astype('float32')

# Create a Logistic Regression model
logistic_regression = LogisticRegression()

# Train the model
logistic_regression.fit(X_train, y_train)

# Predict on training data
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)

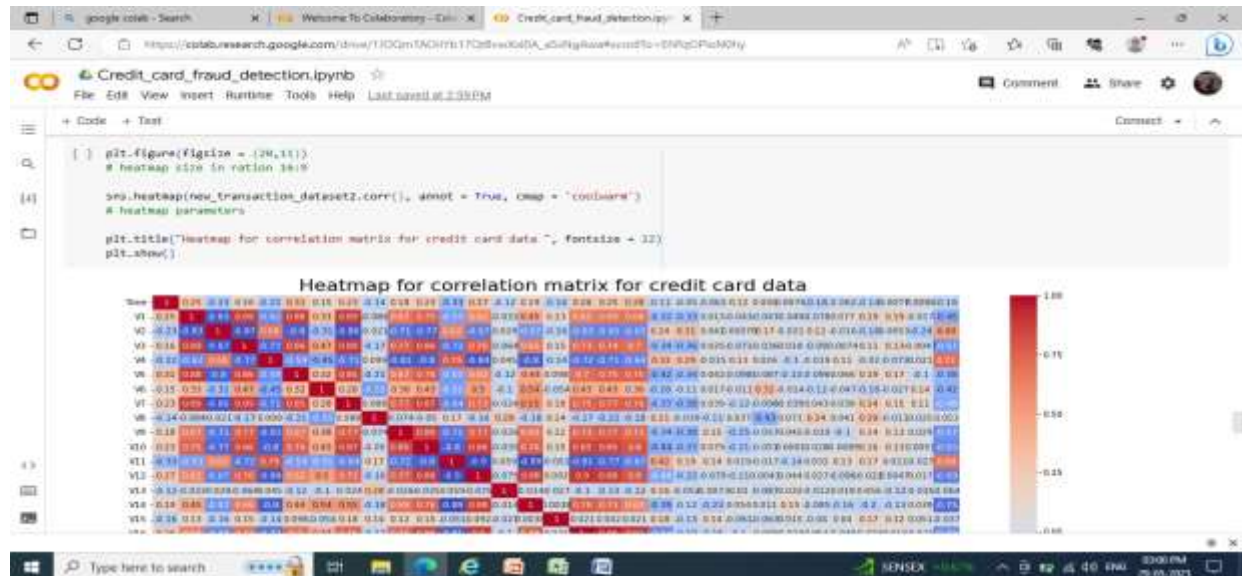
print('Accuracy on Training data : ', training_data_accuracy)

# Predict on test data
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)

print('Accuracy score on Test Data : ', test_data_accuracy)
```

The output of the notebook shows the accuracy on training data as 0.9453621346886912 and the accuracy score on test data as 0.9086294416243854.

## 14.4 HEAT MAP FOR CO-RELATION MATRIX



# **CHAPTER 14**

## **CONCLUSION AND FUTURE ENHANCEMENT**

### **14.1 CONCLUSION**

Machine learning is a potent weapon against credit card fraud because of its capability to scrutinize vast financial data and identify complex fraudulent patterns. Financial card fraud manifests in different forms, from in-person and online scams to identity theft, necessitating sophisticated detection methods.

The journey of using ML for detecting credit card scam involves several stages, including data collection, preprocessing, feature engineering, model training, and evaluation. Different algorithms and models are employed, each tailored to tackle specific fraud challenges.

Success in spotting fraud with ML depends not just on technical proficiency but also on ethical and privacy considerations. Addressing imbalanced data, interpretability of fraud detection models and real-time detection capabilities also contributes to efficient scam detection.

In essence, ML for card fraud detection signifies a substantial transformation in fortifying transactions across various sectors, including financial institutions and credit card companies. By leveraging data-driven insights and advanced algorithms,

businesses and consumers can navigate the digital payment landscape with confidence, knowing they are protected from fraudulent activities.

## 14.2 Future Work

**Feature Engineering:** Explore additional features or transformations of existing features to improve the model's performance. This could involve extracting more information from transaction data or incorporating external data sources.

**Model Selection:** Experiment with other machine learning algorithms such as Gradient Boosting, Support Vector Machines, or Neural Networks to see if they outperform the Random Forest model.

**Imbalanced Data Handling:** Implement techniques to handle class imbalance, such as oversampling minority class instances, under sampling majority class instances, or using algorithmic approaches like SMOTE (Synthetic Minority Over-sampling Technique).

**Hyper parameter Tuning:** Fine-tune the hyper parameters of the chosen model(s) using techniques like grid search or random search to optimize performance.

**Ensemble Methods:** Explore ensemble methods such as stacking or blending multiple models to leverage their collective strengths and improve overall predictive performance.

**Anomaly Detection:** Consider incorporating anomaly detection techniques, such as Isolation Forest or One-Class SVM, to identify unusual patterns in the data that may indicate fraudulent activity.

**Real-time Detection:** Develop a real-time fraud detection system that can analyze transactions as they occur and flag potentially fraudulent ones in real-time, providing immediate alerts to cardholders or financial institutions.

**Interpretability:** Enhance the interpretability of the model by employing techniques such as feature importance analysis or model-agnostic interpretability methods to understand the factors contributing to predictions and build trust in the model.

**Continuous Monitoring and Updating:** Implement a system for continuous monitoring of model performance and regular updating of the model using new data to ensure its effectiveness over time and adapt to evolving fraud patterns.

**Regulatory Compliance:** Ensure that the fraud detection system complies with relevant regulations and standards, such as GDPR or PCI DSS, regarding data privacy and security.

By addressing these aspects, the credit card fraud detection system can become more robust, accurate, and effective in mitigating financial losses due to fraudulent activities.



## **CHAPTER 15**

### **REFERENCE**

1. "Credit Card Fraud Detection using Machine Learning Techniques: A Review" by Shreyasi Bhattacharya and Debanjan Konar.
2. "Credit Card Fraud Detection Using Machine Learning and Data Mining Techniques: A Review" by Fadi Thabtah.
3. "A Comparative Study of Machine Learning Techniques for Credit Card Fraud Detection" by Yanhui Guo and Yijun Sun.
4. "Credit Card Fraud Detection using Machine Learning: A Systematic Review and Meta-Analysis" by Mandeep Kaur and Sarbjeet Singh.
5. "Credit Card Fraud Detection using Machine Learning Techniques: A Review" by Sreeja Nair and Deepa Gupta.
6. "Credit Card Fraud Detection: A Systematic Review and Comparison of Machine Learning Techniques" by Syed Danish Masood and Abdul Basit Dogar.
7. "Machine Learning Techniques for Credit Card Fraud Detection: A Review" by Sonia Jindal and Anupama Marwaha.
8. "A Review on Credit Card Fraud Detection using Machine Learning Techniques" by Ajay Kumar and Rashmi Sinha.
9. "Credit Card Fraud Detection using Machine Learning Techniques: A Review" by Vivek Malik and Akanksha Pandey.
10. "A Survey on Credit Card Fraud Detection Techniques using Machine Learning" by Prashant Kadam and Prof. Manish Soni.
11. "Credit Card Fraud Detection using Machine Learning: A Review" by Manasi Pradhan and Siddhartha Choubey.

12. "Machine Learning Techniques for Credit Card Fraud Detection: A Review" by Vishal Aggarwal and Neha Gupta.
13. "A Comprehensive Review on Credit Card Fraud Detection using Machine Learning Techniques" by Satish Kumar and Kirti Gupta.
14. "Credit Card Fraud Detection using Machine Learning Techniques: A Review" by Pooja Kumari and Dr. Jagdish Lal Raheja.
15. "A Survey of Machine Learning Techniques for Credit Card Fraud Detection" by Vishal Gupta and Dr. Sanjay Silakari.
16. "Credit Card Fraud Detection using Machine Learning: A Review" by Anamika Verma and Dr. S. A. Ali.
17. "A Survey on Machine Learning Techniques for Credit Card Fraud Detection" by Arpit Kumar and Shikha Sharma.
18. "Credit Card Fraud Detection using Machine Learning: A Review" by Ashish Kumar and Dr. Arun Sharma.
19. "A Survey of Credit Card Fraud Detection Techniques using Machine Learning" by Shubham Gupta and Dr. Dinesh Kumar.
20. "Credit Card Fraud Detection using Machine Learning: A Review" by Ruchi Verma and Dr. Sunil Kumar.
21. "A Comparative Study of Machine Learning Algorithms for Credit Card Fraud Detection" by Ravi Kumar and Dr. S. K. Pandey.
22. "Credit Card Fraud Detection using Machine Learning: A Review" by Jyoti Verma and Dr. S. K. Gupta.
23. "A Comprehensive Survey on Credit Card Fraud Detection Techniques using Machine Learning" by Deepak Yadav and Dr. M. S. Gupta.
24. "Credit Card Fraud Detection using Machine Learning: A Review" by Pawan Kumar and Dr. V. K. Singh.