VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELGAVI-590018

Project Report

On

"Credit Card Detection Using Predictive Modelling"

submitted in partial fulfilment of the requirement for the 8th semester

BACHELOR OF ENGINEERING

In

ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

M. Durga Venkata Prasad Reddy	**1RL18EC066**
S. Jagadeesh	**1RL18EC102**
T. Bhanu Prakash Reddy	**1RL18EC113**
V. Pavan Kumar	**1RL18EC118**

UNDER THE GUIDANCE OF

Dr.ShivaPrasad K.M
Professor &
Vice Principal

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

R.L. JALAPPA INSTITUTE OF TECHNOLOGY

KODIGEHALLI, DODDABALLAPUR-561203
2021- 2022

R.L. JALAPPA INSTITUTE OF TECHNOLOGY

Kodigehalli, Doddaballapura, Bangalore-561203

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that the Project Report entitled "**Credit Card Fraud Detection Using Predictive Modelling**" is a bonafide work carried out by **Mallu Durga Venkata Prasad Reddy(1RL18EC066), Sankara Jagadeesh(1RL18EC102), Thalapagala Bhanu Prakash Reddy(1RL18EC113), Vaduguru Pavan Kumar (1RL18EC118)** in fulfilment for the award of the degree of Bachelor of Engineering in **Electronic and Communication Engineering** of the **Visvesvaraya Technological University**, Belagavi during the academic year 2021-2022. It iscertified that all corrections and suggestions indicated for Internal Assessment have been incorporated in the report. The project report has been approved as it satisfies the academicrequirements prescribed for the Bachelor of Engineering degree.

Internal Guide
Dr. ShivaPrasad K.M
Professor &
Vice Principal
Dept. of ECE, RLJIT

HoD
Dr. Anil Kumar .C
Associate Professor
Dept. of ECE, RLJIT

Principal
Dr. Sreenivasa Reddy .M
RLJIT

Project Examiners :

1. ______________________

2. ______________________

ACKNOWLEDGEMENT

The completion of any work is a showcase of constant dedication and co -operation of many people who lent their hands which went seen or unseen.

We would like to express regards to **Dr. Sreenivasa Reddy.M Principal RLJIT** for all his prolific support in all endeavours.

Our sincere thanks to **Dr. Anil Kumar.C Associate Professor & HoD** Department of Electronics and Communication Engineering, RLJIT, for his support and encouragement.

At this outset, we express gratitude to our project guide **Dr. ShivaPrasad K.M Professor & Vice Principal** Department of Electronics and Communication Engineering for his seamless guidance at all stages of this complete exercise.

The actual people behind this whole exercise, who not only organized it very well but also saw to it that best technology reached us, is our **Project Coordinators Dr. Madhu Chandra G Associate Professor and Dr.Anil kumar.C Associate Professor** Department of Electronics and Communication Engineering, RLJIT, thanks to them.

We are thankful to all our professors and the entire department of **ELECTRONICS AND COMMUNICATION ENGINEERING** for their co-operation and suggestion. Without their co-operation and suggestions our course would not have been a successful one.

We also thank to our parents, family members and friends for their continuous encouragement and support.

M. Durga Venkata Prasad Reddy (1RL18EC066)

S. Jagadeesh(1RL18EC102)

T. Bhanu Prakash Reddy (1RL18EC112)

V. Pavan kumar (1RL18EC118)

ABSTRACT

Billions of dollars of loss are caused every year by fraudulent credit card transactions. The design of efficient fraud detection algorithms is key for reducing these losses, and more and more algorithms rely on advanced machine learning techniques to assist fraud investigators. The design of fraud detection algorithms is however particularly challenging due to the non-stationary distribution of the data, the highly unbalanced classes distributions and the availability of few transactions labelled by fraud investigators. At the same time public data are scarcely available for confidentiality issues, leaving unanswered many questions about what is the best strategy.

In this project we aim to provide answers by focusing on crucial issues such as:

i) Why and how under sampling is useful in the presence of class imbalance (i.e. frauds are a small percentage of the transactions),

ii) How to deal with unbalanced and evolving data streams (non-stationarity due to fraud evolution and change of spending behaviour),

iii) How to assess performances in a way which is relevant for detection and

iv) How to use feedbacks provided by investigators on the fraud alerts generated. Finally, we design and assess a prototype of a Fraud Detection System able to meet real-world working conditions and that is able to integrate investigators' feedback to generate accurate alerts.

Key Words—Credit card, Fraud detection, Online shopping, E-commerce, Logistic regression.

CONTENTS

LIST OF FIGURES

ABBREVIATIONS

WORD	ABBREVIATION
AI	Artificial Intelligence
APR	Annual Percentage Rate
ASIC	Australian Securities and Investment Commission
ATM	Automated Teller Machine
AUC	Area Under the Curve
BNPL	Book Now Pay Later
BSD	Berkeley Software Distribution
CART	Classification and Regression Tree
CNP	Card Not Present
DT	Decision Tree
EFTPOS	Electronic Funds Transfer at Point Of Sale
FTC	Federal Trade Commission
GUI	Graphical User Interface
IDE	Integrated Development Environment
KNN	K-Nearest Neighbor
LR	Logistic Regression
ML	Machine Learning
OS	Operating System
PCA	Principal Component Analysis
PIP	Preferred Installer Program
RAM	Random Access Memory
RF	Random Forest
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

The online shopping growing day to day. Credit cards are used for purchasing goods and services with the help of virtual card and physical card whereas virtual card for online transaction and physical card for offline transaction. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. In online payment mode, attackers need only little information for doing fraudulent transaction (secure code, card number, expiration date etc.). In this purchase method, mainly transactions will be done through Internet or telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to analyse the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behaviouristic profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system.

1.1 Objective

The key objective of any credit card fraud detection system is to identify suspicious events and report them to an analyst while letting normal transactions be automatically processed.

For years, financial institutions have been entrusting this task to rule-based systems that employ rule sets written by experts. But now they increasingly turn to a machine learning approach, as it can bring significant improvements to the process.

1 Higher accuracy of fraud detection. Compared to rule-based solutions, machine learning tools have higher precision and return more relevant results as they consider multiple additional factors. This is because ML technologies can consider many more data points, including the tiniest details of behaviour patterns associated with a particular account.

2 Less manual work needed for additional verification. Enhanced accuracy leads reduces the burden on analysts. "People are unable to check all transactions manually, even if we are talking about a small bank," Alexander Konduforov, data science competence leader at AltexSoft, explains. "ML-driven systems filter out, roughly speaking, 99.9 percent of normal patterns leaving only 0.1 percent of events to be verified by experts."

3 Fewer false declines. False declines or false positives happen when a system identifies a legitimate transaction as suspicious and wrongly cancels it.

1.2 Existing Methods

This was on k-means Algorithm implementation, Only the two features with the most variance were used to train the model. The model was set to have 2 clusters, 0 being non-fraud and 1 being fraud. We also experimented with different values for the hyper parameters, but they all produced similar results. Changing the dimensionality of the data (reducing it to more dimensions than 2) also made little difference on the final values.

Disadvantages:

The Clustering doesn't produce the less accuracy when compared to Regression methods in scenarios like credit card fraud detection. Comparatively with other algorithms k-means produce less accurate scores in prediction in this kind of scenarios.

1.3 Proposed Method

Our goal is to implement machine learning model in order to classify, to the highest possible degree of accuracy, credit card fraud from a dataset gathered from Kaggle. After initial data exploration, we knew we would implement a logistic regression model for best accuracy reports.

Logistic regression, as it was a good candidate for binary classification. Python sklearn library was used to implement the project, We used Kaggle datasets for Credit card fraud detection, using pandas to data frame for class ==0 for no fraud and class==1 for fraud, matplotlib for plotting the fraud and non-fraud data, train_test_split for data extraction (Split arrays or matrices into random train and test subsets) and used Logistic Regression machine learning algorithm for fraud detection and print predicting score according to the algorithm. Finally Confusion matrix was plotted on true and predicted.

Advantages:

- The results obtained by the Logistic Regression Algorithm is best compared to any other Algorithms.
- The Accuracy obtained was almost equal to cent percent which proves using of Logistic algorithm gives best results.
- The plots that were plotted according to the proper data that is processed during the implementation.

1.4 Problem Statement

Credit card fraud stands as major problem for word wide financial institutions. Annual lost due to it scales to billions of dollars. We can observe this from many financial reports. Such as (Bhattacharyya et al., 2011) 10th annual online fraud report by Cyber Source shows that estimated loss due to online fraud is $4 billion for 2008 which is 11% increase than $3.6 billion loss in 2007and in 2006, fraud in United Kingdom alone was estimated to be £535 million in 2007 and now costing around 13.9 billion a year (Mahdi et al., 2010). From 2006 to 2008, UK alone has lost £427.0 million to £609.90 million due to credit and debit card fraud (Woolsey &Schulz, 2011). Although, there is some decrease in such losses after implementation of detection and prevention systems by government and bank, card-not-present fraud losses are increasing at higher rate due to online transactions. Worst thing is it is still increasing un-protective and un-detective way.

Over the year, government and banks have implemented some steps to subdue these frauds but along with the evolution of fraud detection and control methods, perpetrators are also evolving.

1.5 Problem Solution

In this project, fraud detection using Machine Learning algorithms is proposed. The proposed system uses logistic regression to build the classifier to prevent frauds in credit card transactions. To handle dirty data and to ensure a high degree of detection accuracy, we followed the data pre-processing steps which involves filling of null data.

CHAPTER 2

LITERATURE SURVEY

The literature survey describes the different work carried out in the field of Credit Card Fraud Detection which is summarized as below:

S.NO	YEAR	AUTHOR	TITLE	OUTCOME
1	2021	D. D. Borse, P. S. H. Patil, and S. Dhotre.	Credit Card Fraud Detection Using Naïve Bayes.	Machine learning's Naive Bayes classification was used to predict common or fraudulent transactions. The accuracy, recall, precision, F1 score, and AUC score of the Naive Bayes classifier are all calculated. Asha R B et al. They have proposed a deep learning-based method for detecting fraud in credit card transactions. Using machine-learning algorithms such as support vector machine, k-nearest neighbour, and artificial neural network to predict the occurrence of fraud. used.
2	2021	G. Kibria and M. Sevkli.	Application of Deep Learning for Credit Card Approval.	Using the grid search technique, create a deep learning model. The built model's performance is compared to the performance of two other traditional machine-learning algorithms: logistic regression (LR) and support vector machine (SVM). The developed model is applied to the credit card data set and the results are compared to logistic regression and support vector machine models.
3	2020	Thabtah, Fadi	Data imbalance in classification: Experimental evaluation	Second, the term “noisy data” refers to the existence of outliers within the data employed for training. Outliers can be seen outside of the normal context of the data.

				This issue also leads to poor detection accuracy , Third, the concept of drift means that the behaviour of the client changes, resulting in changes in the data stream when dealing with online data detection in real time.
4	2020	Gianini, Gabriele	Managing a pool of rules for credit card fraud detection by a Game Theory based approach.	evaluated the current fraud detection system with regard to credit card transactions. The problem is that there are two stages for automatic classification: real-time (RT) and near-real-time (NRT). They focused on the NRT stage by using a rule-based classification technique that considers the final evaluation of the human element of fraud. The authors did not improve the design of the system, discover any new rules, or improve the arithmetic efficiency of individual rules. Instead, they manipulated the rules to form a decision-making system to improve both the accuracy and the performance. The key idea is to calculate the contribution of each rule involved in the system.
5	2020	Janbandhu, Ruchika, Shameedha Begum, and N. Ramasubramanian.	Credit Card Fraud Detection.	it is estimated that 10,000 transactions take place via credit cards every second worldwide. Owing to such a high transaction frequency, credit cards have become the primary targets of fraud. Indeed, since the Diners Club released its first credit card in 1950, credit card companies have been fighting against fraud.

6	2020	S. H. Projects and W. Lovo.	An analysis of fraud detection techniques.	Online payment methods have been used widely as an outcome of the rapid increase in non-cash electronic transactions. Credit cards represent one of the electronic payment methods A credit card is a thin rectangular piece of plastic or metal issued by a bank or financial services company to a consumer (cardholder) to facilitate payment to a merchant of goods and services. It is based on the consumer's promise to the card issuer. The card issuer (usually a bank) opens an account, which is usually circling, and contributes a line of credit to the user. Which the user can use to make a payment. With a card-based payments accounting for approximately 51% of transactions.
7	2020	R. San Miguel Carrasco and M.-A. Sicilia-Urban.	Evaluation of Deep Neural Networks for Reduction of Credit Card Fraud Alerts.	Deep neural networks have been used to test and measure their ability to detect false positives by processing alerts generated by a fraud detection system. Ten neural network architectures classified a set of alerts triggered by an FDS as either valid alerts, representing real fraud cases, or incorrect alerts, representing false positives. When capturing 91.79 percent of fraud cases, optimal configuration achieved an alert reduction rate of 35.16 percent, and a reduction rate of 41.47 percent when capturing 87.75 percent of fraud cases. Kibria and Sevkli.
8	2020	X. Yu, X. Li, Y. Dong, and R. Zheng.	A Deep Neural	They have proposed a deep network algorithm for fraud detection A deep neural network algorithm for detecting credit card

			Network Algorithm for Detecting Credit Card Fraud.	fraud was described in the paper. It has described the neural network algorithm approach as well as deep neural network applications. The pre-processing methods and focal loss; for resolving data skew issues in the dataset. Siddhant. Bagga et al.
9	2019	Yousefi, Niloofar, Marie Alaghband, and Ivan Garibay	Comprehensive Survey on Machine Learning Techniques and User Authentication Approaches for Credit Card Fraud Detection.	The aim of fraud is to achieve personal or financial gain through deception. Based on this, fraud detection and prevention are the two significant methods for avoiding loss due to fraud. Fraud prevention is the proactive technique for avoiding the occurrence of fraudulent acts, and fraud detection is the technique for the detection of fraudulent transactions by fraudsters.
10	2019	S. Mittal and S. Tyagi.	Performance evaluation of machine learning algorithms for credit card fraud detection.	To evaluate the underlying problems, some popular machine learning-algorithms in the supervised and unsupervised categories were selected. A range of supervised learning algorithms, from classical to modern, have been considered. These include tree-based algorithms, classical and deep neural networks, hybrid algorithms and Bayesian approaches.
11	2018	V. Patil and U. Kumar Lilhore.	A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection.	A supervised learning methodology, graphical representation of possible solutions to a choice based on certain situations and it is a tree-structured classifier. It starts with a root node where inside nodes represent the features of a dataset, branches symbolize the decision rules and each leaf node represents the

				result. In a decision tree and they have the purposes of deciding and communicating respectively.
12	2018	R. R. Popat and J. Chaudhary.	A Survey on Credit Card Fraud Detection Using Machine Learning.	Supervised algorithms were presented Deep learning, Logistic Regression, Nave Bayesian, Support Vector Machine (SVM), Neural Network, Artificial Immune System, K Nearest Neighbour, Data Mining, Decision Tree, Fuzzy logic based System, and Genetic Algorithm are some of the techniques used. Credit card fraud detection algorithms identify transactions that have a high probability of being fraudulent. We compared machine-learning algorithms to prediction, clustering, and outlier detection.
13	2018	S. Xuan, G. Liu, Z. Li, L. Zheng, S.Wang, and C. Jiang.	Random forest for credit card fraud detection.	For training the behavioural characteristics of credit card transactions, the Random Forest classifier was used. The following types are used to train the normal and fraudulent behaviour features Random forest-based on random trees and random forest based on CART. To assess the model's effectiveness, performance measures are computed.
14	2018	Vimala Devi and K. S. Kavitha	Fraud Detection in Credit Card Transactions by using Classification Algorithms	To detect counterfeit transactions, three machine-learning algorithms were presented and implemented. There are many measures used to evaluate the performance of classifiers or predictors, such as the Vector Machine, Random Forest, and Decision Tree. These metrics

				are either prevalence-dependent or prevalence-independent. Furthermore, these techniques are used in credit card fraud detection mechanisms, and the results of these algorithms have been compared.
15	2017	P. Save, P.Tiwarekar, K. N., and N. Mahyavanshi.	A Novel Idea for Credit Card Fraud Detection using Decision Tree.	proposed a model based on a decision tree and a combination of Luhn's and Hunt's algorithms. Luhn's algorithm is used to determine whether an incoming transaction is fraudulent or not. It validates credit card numbers via the input, which is the credit card number. Address Mismatch and Degree of Outlier Ness are used to assess the deviation of each incoming transaction from the cardholder's normal profile. In the final step, the general belief is strengthened or weakened using Bayes Theorem, followed by recombination of the calculated probability with the initial belief of fraud using an advanced combination heuristic.

Summary of Literature review :

The detailed literature review summarized that their exist various techniques for fraud detection of real time transaction. The fraud detection accuracy depends on the Machine Learning algorithms like Naïve Bayes, Deep Neural Networks, Supervised and Unsupervised learning algorithms.

CHAPTER 3

OVERVIEW OF CREDIT CARDS

3.1 Credit Card

A credit card is a type of credit facility, provided by banks that allow customers to borrow funds within a pre-approved credit limit. It enables customers to make purchase transactions on goods and services. The credit card limit is determined by the credit card issuer based on factors such as income and credit score, which also decides the credit limit.

The credit card information includes credit card number, cardholder’s name, expiration date, signature, CVC code, etc. The best part about a credit card is that it is not linked to a bank account. So, whenever you swipe your credit card, the amount is deducted from your credit card limit, not your bank account. You can use it to pay for food, clothes, take care of medical expenses, travel expenses, and other lifestyle products and emergency services.

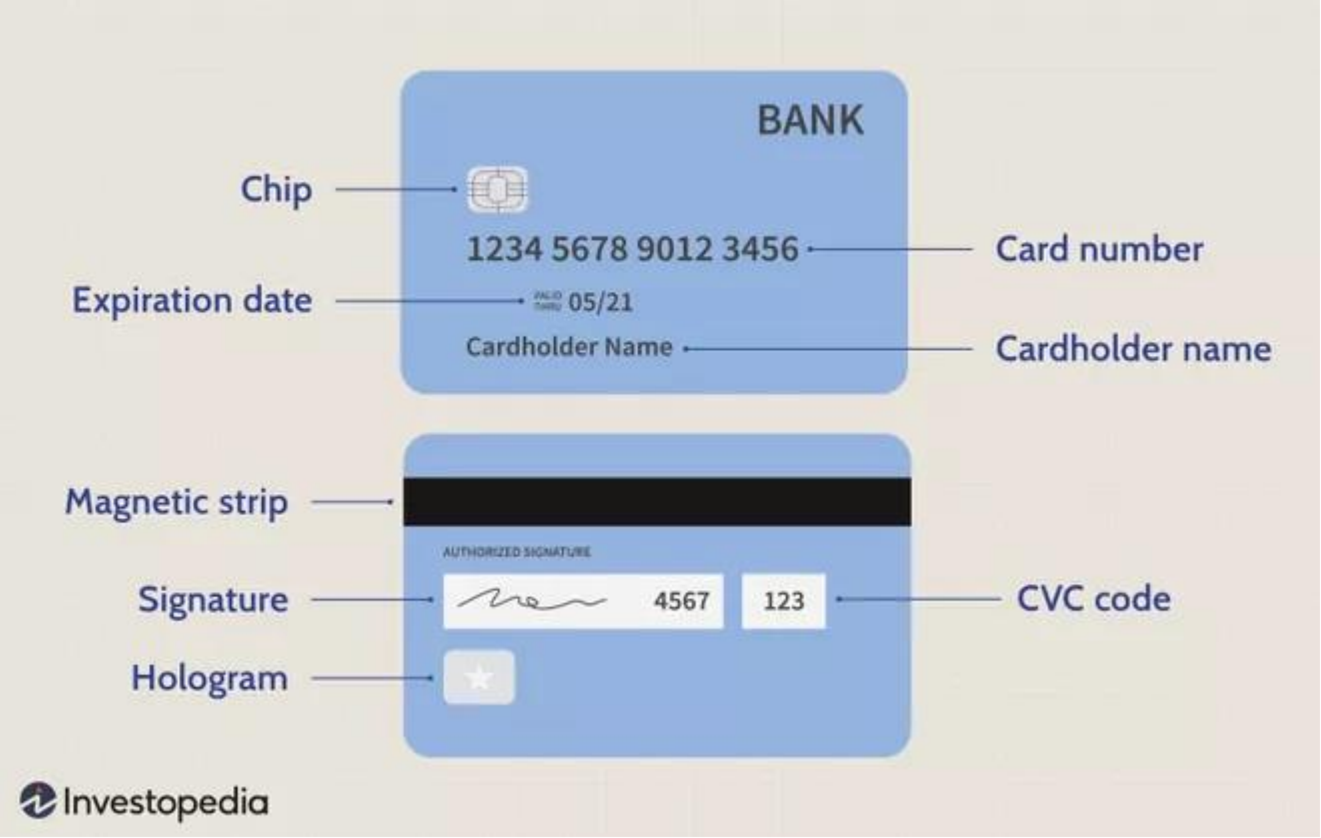


Fig 1 : External view of a Credit Card

3.2 Significance of Credit Card

You can find almost all banks and financial institutions offering various types of credit cards. They are accepted as a means of payment at any place that offers you goods or service. The uses of credit cards are extended to buying groceries, clothing and accessories, booking a movie ticket, shopping online, buying home appliances, paying your utility and mobile bills, and many more.

The benefits of credit cards are innumerable, and some prime ones are:

3.2.1 Buy on credit:

What makes a credit card attractive is the credit limit allowed to the cardholder. You can buy anything within that limit and pay later. Your monthly budget will not affect, even if you buy items of high value on credit. One among the most important benefits of credit card is you can convert the total amount of your purchases into low-cost EMIs to enable you to repay it easily over a period of time. This has helped revolutionize the shopping experience.

3.2.2 Most accepted method of payment:

You can travel anywhere, without carrying much money if you have this card. Being the most accepted method of payment, you can use a credit card to pay anything.

3.2.3 Interest-free cash withdrawals:

There are a few credit cards that allow you to withdraw money up to a certain limit in case of emergency, with no interest charged up to 45 to 50 days. You can make use of it in times of financial emergency.

3.2.4 Unlimited reward points:

These cards come with reward points when you use them. For instance, IDFC FIRST Bank credit card offer unlimited and never-expiring reward points, which are easily redeemable.

3.2.5 Insurance coverage

You get personal accident coverage, as well as comprehensive travel insurance coverage and this is one of the significant benefits of credit cards, which make them attractive.

3.2.6 Make travel easy:

The uses of credit cards in travel make them important. When it comes to IDFC FIRST Bank credit cards, they give you a unique experience through complimentary lounge access at the airports and railway stations in India and priority check-in. Other than these, you can also enjoy discounts on food in more than 280 restaurants.

3.2.7 Discounts and cashbacks:

The advantages of credit cards extend to discounts on some of your favourite entertainment and dining outlets, travel and shopping apps, etc. You can also enjoy fuel surcharge waivers at petrol pumps across the country. To check the host of offers.

3.2.8 Improve your credit score:

The benefits of credit cards do not limit to shopping on credit; instead, it helps improve your credit score. If you know how to use a credit card and how to make use of the credit period, and repay the amount used on time, you can boost your CIBIL score. This will help you obtain loans, without any difficulty in future.

3.2.9 Offers safety:

You don’t have to carry much money if you have a credit card.

3.2.10 Keep track of your expenses:

The statements you get every month from net banking helps you check your expenses and plan the repayment without any delay.

3.2.11 EMI Payments :

Among a distinct credit card advantage is the fact that a credit card can be used for buying flagship items at affordable EMIs. These can be repaid over a duration you can select.

3.2.12 Easy Approval :

A credit card is available online and offline. Furthermore, eligibility criteria are easy to fulfil, and basic documentation is required.

3.2.13 Customised Card Limit :

The credit card limit varies from one card to another and is largely based on the discretion of the card issuer. You also have to stipulated credit score, and the higher this is, the higher your credit limit will be.

3.2.14 Loans During Emergencies :

Credit card facilities may be used to get a personal loan in case of emergency expenses.

3.2.15 Cash Withdrawal :

Another advantage of a credit card over a debit card is that you can also withdraw advance cash from ATMs with the facility of repaying amounts when you have to settle your bill.

3.2.16 Payment Security :

A credit card is a digital instrument that offers you safety in payments. With multi-factor authentication and in-hand security features, you needn’t have cause to be concerned.

3.3 History of Credit Card

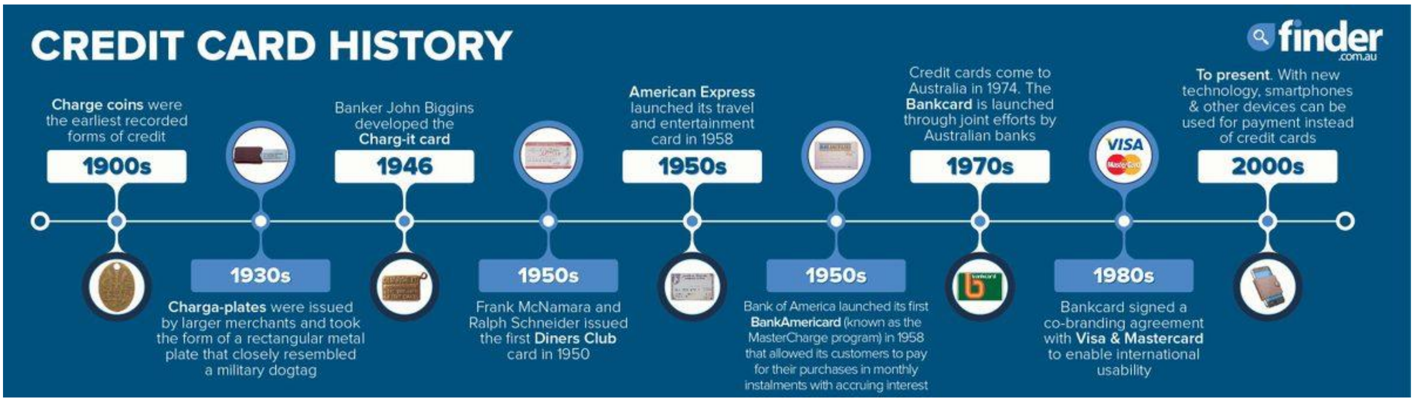

CREDIT CARD HISTORY
finder.com.au
Charge coins were the earliest recorded forms of credit
1900s
1930s
Charga-plates were issued by larger merchants and took the form of a rectangular metal plate that closely resembled a military dogtag
Banker John Biggins developed the Charg-it card
1946
1950s
Frank McNamara and Ralph Schneider issued the first Diners Club card in 1950
American Express launched its travel and entertainment card in 1958
1950s
1950s
Bank of America launched its first BankAmericard (known as the MasterCharge program) in 1958 that allowed its customers to pay for their purchases in monthly instalments with accruing interest
Credit cards come to Australia in 1974. The Bankcard is launched through joint efforts by Australian banks
1970s
1980s
Bankcard signed a co-branding agreement with Visa & Mastercard to enable international usability
To present. With new technology, smartphones & other devices can be used for payment instead of credit cards
2000s

Fig 2. History of Credit Card

Timeline of credit card history

While forms of credit can be traced back to the dawn of civilisation, we've started at the beginning of the 20th century. This was when synthetic plastics (which helped give credit cards their nickname) were invented.

3.3.1 1900s to 1930s: Early forms of credit

1. *1900s*

Carried over from the late 19th century, charge coins were the earliest recorded forms of credit. Large merchants, such as hotels, department stores and petrol companies, would issue their regular customers with charge coins to use with their store charge accounts.

These coins took all shapes and forms and were made out of metal or celluloid, bearing the customer’s charge account number and the merchant’s name and logo for easy imprinting on sales slips. Since there were no other identifying marks on the charge coins, they could be easily stolen and used for fraud.

2. *1930s*

Charga-plates came next, taking the form of a rectangular metal plate that closely resembled a military dog tag.The Charga-plate bore the customer’s name, address, account number and sometimes their signature, which was more helpful for preventing fraud. But this form of credit was still only issued by large merchants and could only be used in the issuing store.

3.2.2 1940s to 1960s: Charge cards and credit cards

1. *1940s*

In 1946, banker John Biggins developed the Charga-it card, which was the first attempt by a bank to issue a card that customers could use at more than one merchant store.

Issued by the Flatbush National Bank of Brooklyn in New York, customers could use this card at a range of merchants within the state and the bank would settle payments with merchants on their behalf.

2. *1950s*

About 4 years after that, Frank McNamara and Ralph Schneider issued the first Diners Club card in 1950. These cards were made of cardboard and could be used in participating restaurants. Cardholders paid an annual fee of US$3 while restaurants paid 7% on transaction values. This was still a charge card and not a credit card per se, as it didn't offer revolving credit.

3. *1958*

American Express launched its travel and entertainment card in competition with the Diners Club card, while Bank of America launched its first BankAmericard (later renamed Visa in 1976) in Fresno California.

While Diners Club and American Express issued charge cards, Bank of America issued credit cards that allowed its customers to pay for their purchases in monthly instalments with accruing interest (or “carrying charges”).

4. *1960s*

Around 1965, Bank of America started licensing its Californian credit card system to banks across America. This move also led to the formation of a national bank card association for enabling nationwide use of the BankAmericard (often seen as the beginnings of Visa).

In 1966, Master Charge (renamed Mastercard in 1979) jumped on the scene as a cooperative of North-eastern banks wishing to honour cards issued by one another. These 2 card schemes have gone head-to-head ever since.

3.3.3 1970s: Credit cards come to Australia

1. *1970s*

Before 1974, only store-issued cards (which could be used exclusively in the issuing store) were used in Australia, with a small number of Diners Club and American Express credit cards accessible to the wealthy.

2. *1974*

The Bankcard was launched through the joint effort by Australian banks. Working together, they had developed their own card network and implemented the technology needed for a nationwide shared facility. Each bank issued its own bank-branded Bankcard and enforced its own card regulations and customer relations.

3. *1976*

By 1976, the Bankcard was a hit. There were 1,054,000 Bankcard holders and almost 49,000 participating merchants. The first ATMs began popping up in Australia in 1977, and by 1978, Bankcards could be used across the nation.

3.3.4 1980s to 1990s: The global credit card market

1. *1980*

In 1980, Bankcard signed a co-branding agreement with Visa and Mastercard to enable international usability. But this was a false start, with that agreement quickly dissolving in the same year.

2. *1983-84*

The EFTPOS system was launched in Australia, and Bankcard was successfully introduced in New Zealand. As a result, there were over 5 million cards in circulation by 1984.

At the same time, both Visa and Mastercard were launched internationally. This saw the rise of international credit cards in Australia and the beginning of Bankcard's decline.

3. *1990s*

By 1994, the number of Bankcards in circulation was down to 3.9 million – a trend that would lead to its subsequent end in 2007.

But credit cards kept growing in popularity, helped by the invention of telephone and Internet banking in the mid-1990s and BPAY in 1997.

3.3.5 2000s: The rise of rewards, perks and fraud

1. *2000s*

The start of the 21st century saw the rise of the rewards credit card as people with credit cards looked for ways to get more bang for their buck.

Earning points per $1 spent and introductory bonus point offers gave people a way to get frequent flyer rewards faster, while credit card reward programs offered flexibility for travel, shopping and cash rewards. Perks like airport lounge access, complimentary travel insurance and travel or flight credit have also become popular.

2. *Mid-2000s*

Innovations such as contactless payments, mobile banking and Internet shopping have led to higher incidences of credit card fraud and scams. In particular, phishing, skimming and hacking all became significant threats that took the focus off old-school physical theft.

As a result, card companies, banks and governments have all created policies, services and tools to help keep accounts and personal information safe – including zero-liability protection, fraud monitoring services, government support services and federal investigations into criminal card activity.

3.3.6 2010s: Credit card reforms, cardless payments and After pay

1. *2010*

The Australian Securities and Investment Commission (ASIC) became the sole regulator for credit law, following the National Consumer Credit Protection Act (2009). This marked the start of a series of updates to credit card regulations, including the following:

- Key Fact Sheets for all cards
- Restricting unsolicited offers for credit or credit limit increases
- Repayment allocation requirements
- Credit limit requirements
- The option to reduce your credit limit or cancel a card online

2. *2016*

Australians were quick to adopt tap-and-pay technology built into credit cards and debit cards. In 2016, Australians were the global leaders on contactless payments according to research firm RFi Group, which had found 59% of all Aussies had made a contactless payment.

But cards were still king, with RFi research at the time showing only 10% of Australians said they used a mobile wallet and only 24% would consider it in the future.

3. *2018-19*

As newer smartphones came onto the market and mobile wallet apps like Apple Pay, Google Pay and Samsung Pay became more popular, Australia saw a surge in mobile payment adoption. By 2019, contactless payments from smartphones, smartwatches and other mobile devices made up 8% of all in-person card payments, up from just 1% in 2016 according to data from the Reserve Bank of Australia.

Mobile and contactless payments have continued to grow as apps make it easier to pay both in-person and online (and following initial health concerns around card payments when the coronavirus pandemic began).

4. *2018*

After pay was launched in Australia and kickstarted the buy now pay later (BNPL) trend. At the time, BNPL was not seen as a direct competitor to credit cards, but now it's common to see them compared side by side.

3.3.7 2020s: Current credit card trends

At the start of 2022, around 68% of Australian adults had a credit card, with more than 13 million credit cards in circulation and a national debt of $18.5 billion accruing interest (you can read more stats here). Current card trends include the following:

- **No interest credit cards.** Launched in 2020, these cards don't charge interest and have a flat monthly fee when you use them or carry a balance. No interest credit cards are designed as an alternative to buy now pay later and combine features of both BNPL and traditional credit cards.
- **Virtual credit cards.** The widespread use of mobile wallets has seen plastic credit cards make way for virtual ones. While it's common to have a plastic card as well as a digital one, some credit card providers are starting to offer instant virtual and virtual-only options, such as the PayPal Rewards Card and MONEYME Freestyle Virtual Card.
- **Introductory offers.** Bonus points and 0% interest offers have been popular for a long time, but a more recent trend is one-off cashback offers alongside other credit card perks. Overall interest in cashback has also grown and in 2021, American Express launched its Cashback Credit Card – the only dedicated cashback rewards card you can currently get with a major brand in Australia.

- **Credit card debt.** Since around 2020, credit card debt has followed a downward trend. Finder analysis has shown that the average credit card balance in January 2022 was $2,721, compared to $3,392 in December 2019.

3.4 Types of Credit Cards

Following are the different types of credit cards outlined each type of card might work for your spending and your financial goals.

3.4.1 Rewards credit cards

Rewards credit cards typically give you points or cash back based on a percentage of your spending and some even offer bonus points in popular categories like groceries, gas and dining out.

Rewards credit cards also tend to offer at least a few different ways to redeem your points, often including options for statement credits, gift cards or merchandise. This makes them a great credit card option for everyday expenses when you know you can pay off your card right away. By using a rewards credit card to cover your basic purchases like groceries and household supplies, you can earn cash back and travel rewards for purchases you needed to make anyway.

3.4.2 Cash back credit cards

Cash back credit cards make it easy for you to earn cash back or statement credits on your spending, although how rewards are doled out varies from card to card. Some options in this niche offer a flat rate of rewards while others offer bonus points in certain categories like dining or travel. Some even offer bonus rewards in categories that change each quarter, as well as a flat rate of rewards on all non-bonus purchases.

Many cash back credit cards come with no annual fee as well, although some with more generous bonus offers and rewards schemes charge annual fees on the modest side, usually under $100.If you tend to spend more in particular categories, like groceries or dining, you might choose a bonus category card rather than a flat rate card, which is ideal for those with varied spending looking for an everyday card.

3.4.3 Travel credit cards

Travel credit cards offer you the opportunity to earn rewards that are geared specifically toward travel, whether that means earning flexible travel credits you can use toward any travel purchase or even points you can transfer to airline or hotel programs. Some travel credit cards also let you earn points within a specific program, such as a frequent flyer program or hotel loyalty program.

If you often travel for business or pleasure, you can also keep your eye out for luxury travel credit cards that offer perks like airport lounge access, annual travel credits and credits for Global Entry or TSA Precheck. You don’t even have to leave town to start earning points and miles that can make your next trip more affordable—many of the best travel rewards cards let you earn these perks by making everyday non-travel-related purchases.

3.4.4 Balance transfer credit cards

If you have some high-interest credit card debt on your hands, you may be considering using a balance transfer credit card to help manage and pay down that debt. The best balance transfer cards let you secure an introductory 0 percent APR for a period typically between 15 and 21 months, which can give you a nice break from paying interest charges while you focus on paying down your debt. Most cards require you to pay an upfront balance transfer fee of 3 percent or 5 percent, although there are some credit cards with no balance transfer fee. Still, even after taking the balance transfer fee into account, you could save a significant sum of money on interest during your card’s introductory APR offer.

3.4.5 Zero percent intro APR credit cards

You’ll also find an array of credit cards that offer 0 percent intro APR on purchases for a limited time, usually up to 18 months. Credit cards in this niche can be a boon if you need to make a large purchase and want to pay it back over time without interest. On the same token, you’ll also find low-interest credit cards that offer lower than average rates overall and not just during an introductory offer phase.

3.4.6 Business credit cards

Business credit cards allow cardholders to keep their personal and business expenses separate while they earn rewards on all their business spending. Interestingly enough, business credit cards can also be cash back credit cards, general rewards credit cards, travel credit cards or even secured credit cards. You do need to have a business or income-producing activity to qualify for a business credit card.

In general, the signs of a good business credit card are if it helps you benefit from your everyday spending and makes running your business easier. Ideally, you’ll want to find a credit card with a generous rewards program, expense tracking abilities and features that help boost your bottom line. Some business credit cards will give you a flat rewards rate for all of your purchases, whereas other cards reward common business expenses like travel or internet service at higher rates. When it comes to redeeming your rewards, business credit cards generally let you exchange your rewards for either cash back or airline miles.

3.4.7 Student credit cards

Student credit cards are “starter credit cards” of sorts specifically geared to young people with a limited credit history. In other words, application requirements aren’t as stringent, so it’s easier to get approved. Most student credit cards don’t charge an annual fee and many offer bonus perks for good grades as well as rewards for each dollar you spend. If used responsibly, signing up for a student credit card can help young people build their credit and start creating good financial habits.

3.4.8 Secured credit cards

Most credit cards are unsecured, meaning you don’t have to put down any collateral. With secured credit cards, on the other hand, you’re required to put down a cash deposit in order to secure a small line of credit, usually for a similar amount. For example, you might sign up for a secured credit card and put down a $500 initial deposit in order to receive a $500 line of credit. The one-time deposit (and therefore credit limit) can be as low as $49.

While putting down collateral may not seem ideal, secured credit cards are the easiest type of credit card to get approved for, so they are often helpful when you need to build credit from scratch or want to repair your credit after a financial hurdle.

3.4.9 Store credit cards

Store credit cards are offered through retail stores to let consumers charge their purchases and pay them off over time. Store credit cards are generally only used within the specific store that offers them, although some store credit cards can be used within a specific family of stores.

Generally, store-branded credit cards have higher interest rates than general-purpose cards, and they are often more likely to charge deferred interest. Deferred interest means you’ll get a low or 0 percent introductory rate for a period of time, but if you don’t pay the full amount off within that time, you’ll be charged retroactive interest. That being said, if you can pay off your store credit card on time, you may be able to take advantage of some great perks and rewards programs.

3.4.10 Co-branded credit cards

Co-branded credit cards are store or brand credit cards offered through traditional card issuers like Chase, Citi, or American Express. These can include airline credit cards that let you earn miles within a specific frequent flyer program or hotel credit cards that let you earn points within a hotel loyalty program. Some co-branded credit cards also partner with retail stores, although you can typically use them for non-store purchases as well.

Generally, the rewards offered by co-branded credit cards are limited to one brand, but their rewards are solid and in many cases, the value of these rewards (like free hotel nights) end up being worth more than cash back.

3.5 Types of Frauds we can expect in Credit Card

According to the Australian Payments Network, following five are the key types of credit card fraud:

- Card-not-present (CNP) fraud
- Counterfeit and skimming fraud
- Lost and stolen card fraud
- Card-never arrived-fraud
- False application fraud

3.5.1 Card-not-present (CNP) fraud

‘Card not present’ fraud occurs without the use of the physical card, mainly online or over the phone.

Amount lost in 2018: **$477,920,701**

Card-not-present transactions are becoming more and more popular as customers turn away from using their physical cards and simply enter their details to make a purchase. For example, an online purchase made ordering something on eBay is a CNP purchase, even if you’ve still read your details off the card. To be considered ‘card-present’, the card details have to be captured at the point of sale, like being pressed into a contactless reader, inserted into a merchant’s terminal or used at an ATM.

Card-not-present fraud is the biggest contributor to overall credit card fraud, accounting for 85% of all fraud on Australian cards (this also includes debit cards). It increased in size by nearly 8% over June 2017-18 and occurs mainly when credit card details are stolen to make purchases. AusPayNet CEO Dr

Leila Fourie says CNP fraud has become so popular now due to both the growth of eCommerce and the increased security around other types of fraud.

“Combatting CNP fraud is now a key priority across the entire e-commerce community and we’re delighted with the strong progress made this year on a framework for mitigating CNP fraud. We expect this whole of industry approach will help reduce CNP fraud, in the same way chip technology is tackling skimming fraud,” said Dr Fourie.

“Malware and phishing attacks are becoming increasingly sophisticated, so treat unsolicited emails and text messages from people you don’t know with suspicion. Don’t click on the link provided and don’t be tricked into divulging confidential data such as your password.”

So be careful with those credit card details online and don’t speak too loudly when reading them out over the phone. In fact, just be careful about who you allow to read the back of your card in general, since thieves are having an absolute field day with our details online. Once they have your card details they may be able to spend to their heart’s content until:

- They hit your credit limit
- Your account runs out of money
- You contact your bank and tell them to cancel the card as soon as possible
- Some credit card providers can detect suspicious activity on your credit card (e.g. a few multi-thousand-dollar transactions are suddenly being made in Lagos) and may temporarily suspend the card until you confirm whether the activity is really you. This can be a pain when using the card while you’re travelling overseas, but so long as you inform your credit card provider of your travel plans beforehand, you shouldn’t trigger any unnecessary card suspensions.

3.5.2 Counterfeit and skimming fraud

Counterfeit and skimming frauds are those that occur when details are illegally taken to create a counterfeit credit card.

Amount lost in 2018: $14,935,409 (*Source: AusPayNet)*

‘Skimming’ is when a device steals the details of your credit card from its magnetic stripe and commonly occurs when a device, known as a credit card skimmer, is attached to either an ATM or a merchant’s terminal. Skimming can also occur when someone brushes past you with a credit card skimmer. Details obtained via skimming can then be used to create a counterfeit card or to take part in some good old card-not-present fraud.

That near $15 million figure ($23 million when you take overseas fraud into account) might seem like a lot but has actually fallen quite significantly in recent years thanks to increasingly advanced protection offered by chip technology. Over the 17-18 financial year, skimming fraud fell from $42.3 million to $23 million – a record low – and only accounts for 4% of all card fraud now. This is a credit to Australian chip-protection technology, which is among the best in the world. But in America, there are 60 million compromised credit cards and three-quarters of these are estimated to be due to skimming and POS (point-of-sale) breaches, according to Gemini Advisory.

That doesn't mean you shouldn't be careful in Australia though. Keep your card well-within the confines of your wallet or purse, and if an ATM looks like it's been tampered with, report it and move on.

Skimming is also different from phishing, which is when you hand your card details over to someone under the guise of someone or something else. For example, a common phishing scam is when someone creates a fake company that looks like a real one **(**let's say Comonwealth Bank instead of Commonwealth Bank) and sends an email asking for card details. These phishers will often have extremely similar (or even the same) logos as existing companies with similar URLs to boot, so they can be easy to fall victim to.

According to Scamwattch, phishing is the most common kind of fraud in the country, with just under 25,000 reports occurring in 2018.

3.5.3 Lost and stolen card fraud

Lost and stolen card fraud occurs on cards that have been lost or stolen.

Amount lost in 2018: $47,478,058 (*Source: AusPayNet)*

This one should be pretty self-explanatory – if your card has been lost or stolen by a pickpocket, then they are free to use that card until it's cancelled, suspended or it has hit the credit limit. Nearly $50 million was lost to stolen cards from June 2017-18, so know your card's whereabouts at all times. You can avoid the worst of the damage (or all of it) by cancelling or freezing the card as soon as you can by calling your bank. Some of them even let you do this with the click of a button in their mobile banking apps.

If you decide you don't want a credit card anymore, don't just hurl it into the trash. Thieves can still take it and use it since it'll still be active. Cancel it, and then cut it up to avoid having your card stolen from the bin.

3.5.4 Card-never arrived-fraud

'Card never arrived' fraud occurs on cards ordered by a customer that they never receive.

Amount lost in 2018: $6,231,308 (*Source: AusPayNet)*

When you make an application for a credit card, 99% of the time that card will be sent to you in the mail. Card-never-arrived fraud is what happens when that card is either intercepted before it arrives, or if your card thief simply pinched it from your letterbox, which is more likely.

To protect against this type of fraud, the Australian Payments Network recommends installing a lockable mailbox, or at the very least checking your mailbox regularly.

3.5.5 False application fraud

False application fraud occurs where the account was established using someone else's identity or information.

Amount lost in 2018: **$2,393,902 (*Source: AusPayNet)***

Application fraud can be a bunch of different things. It might be that someone applies for a credit card in your name and runs up a bunch of debt, ruining your credit rating. Or maybe they apply for a card in a different name but link your bank account to the card, so you get slugged with the repayments. Someone could run up thousands of dollars on a credit card or completely tarnish your credit score before you realise you've been had.

Back in 2014, an analysis of credit card applications by Veda found $1.6 billion worth of applications for credit were red-flagged as potentially fraudulent. Most credit card providers take application fraud very seriously and have a string of checks and balances to make sure this doesn't happen, but that doesn't mean the occasional fraudster doesn't slip through the proverbial cracks. Make sure you keep track of your bank accounts, keep sensitive information hidden and most importantly, take any kind of fraudulent activity seriously and report is as soon as you can.

3.6 Types of Databases available for Credit Card Fraud Detection

The dataset we're going to use can be downloaded from Kaggle. It contains data about credit card transactions that occurred during a period of two days, with 492 frauds out of 284,807 transactions.

dataset URL : https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

All variables in the dataset are numerical. The data has been transformed using PCA transformation(s) due to privacy reasons. The two features that haven't been changed are Time and Amount. Time contains the seconds elapsed between each transaction and the first transaction in the dataset.

"The datasets contains transactions made by credit cards in September 2013 by European cardholders. This dataset present transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.
It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise."

3.7 Previous methods used Fraud Detection

Earlier methods used for fraud detection are:

3.7.1 Decision Tree

The decision tree method works by using a similarity tree which is created by using decision tree logic. A similarity tree is outlined with nodes and leaves which have attributes and factors. These define the ratio in terms of transactions that satisfy certain conditions. This method is easy to comprehend and display. On the downside, it can be that every transaction needs to be checked individually. This method has been used to provide very good results for several years.

3.7.2 Genetic Algorithms and A Range of Additional Algorithms

Algorithms can be used to detect fraud by using predictive methods. What the algorithms do is establish a set of rules based on logic. This allows the data to be categorized into either non-suspicious or suspicious activity. This credit card fraud detection method has delivered results and is also useful for home insurance data. It is an efficient method when tackling credit card fraud and uses a range of methods that highlight suspicious transactions.

3.7.3 Clustering Techniques

Clustering techniques can be used to detect behavioral fraud. One clustering method is Peer Group Analysis. This is a method that identifies accounts that are behaving in a different way to other accounts. If an account is suddenly behaving differently to previously then this method allows it be flagged. Once flagged, the appropriate methods can be used to contact the customer or block the account to prevent any further fraud taking place.

It can often be the case that a customer is genuinely wishing to make a high-dollar transaction which is unusual to their normal pattern of small purchases. If all is well, then the account will be unblocked.

3.7.4 Neural Networks

Neural networks are also seen as an effective way to combat credit card fraud. The disadvantage of this method is that the method uses data clustering which can only be collated by account type.

3.7.5 Naive Bayes Classifiers

Naive Bayes is a supervised machine learning method developed by John and Langley in 1995. The method uses a dataset with target classes that are known in order to make predictions of future instances. Experiments that have been performed on this method show that it performs well.

3.7.6 K-Nearest Neighbor Algorithms

The K-Nearest Neighbor Algorithm or KNN is a method that uses available instances and then classifies new instances based on similarity. KNN has been used to perform pattern recognition and statistical estimation since the 1970s. The KNN is an instance-based learning method. It begins with a set of instances and compares new instances to the original instances. The K-Nearest Neighbor Algorithm was introduced in 1991 by Aha, Kibler and Albert. This method does have its downfalls as irrelevant attributes can lead to impracticalities and inefficiency.

3.7.7 Support Vector Machines (SVMs)

The Support Vector Machine is a statistical learning method that is useful in credit card fraud detection. If the test instance is within the learned region it will be classed as normal and if it is outside of this region it will be classed as anomalous.

3.7.8 Bagging Ensemble Classifier

Introduced by Leo Breiman in 1994 this method was designed to improve upon machine learning algorithms. It has become popular due to its simple implementation as well as increased accuracy. The bagging ensemble classifier is fast and can handle large databases.

3.8 Best ML Algorithms for applications

3.8.1 Logistic Regression

Logistic regression is foremost used to model a binary (0,1) variable based on one or more other variables, called predictors. The binary variable being modelled is generally referred to as the response variable or the dependent variable. For a model to fit the data well, it is assumed that,

a. The predictors are uncorrelated with one another.
b. That they are significantly related to the response.
c. That the observations or data elements of a model are also uncorrelated.

In logistic regression, the response is binary (0,1) and follows a Bernoulli probability distribution. Since the Bernoulli distribution is a subset of the more general binomial distribution, logistic regression is recognized as a member of the binomial family of regression models. Logistic regression is particularly valuable in that the predictions made from a fitted model are probabilities, constrained to be within the range of values 0–1. More accurately, a logistic regression model predicts the probability that the response has a value of 1 given a specific set of predictor values. Interpretation of logistic model coefficients usually involves their exponentiation, which allows them to be understood as odds ratios. This capability is unique to the class of logistic models, whether observation-based format or in grouped format. The fact that a logistic model can be used to assess the odds ratio of predictors, and also can be used to determine the probability of the response occurring based on specific predictor values, called covariate patterns, is the prime reason it has enjoyed such popularity in the statistical community for the past several decades. Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.

Accuracy Score of Credit Card Fraud Analysis using Logistic Regression is : 93.90862944162437

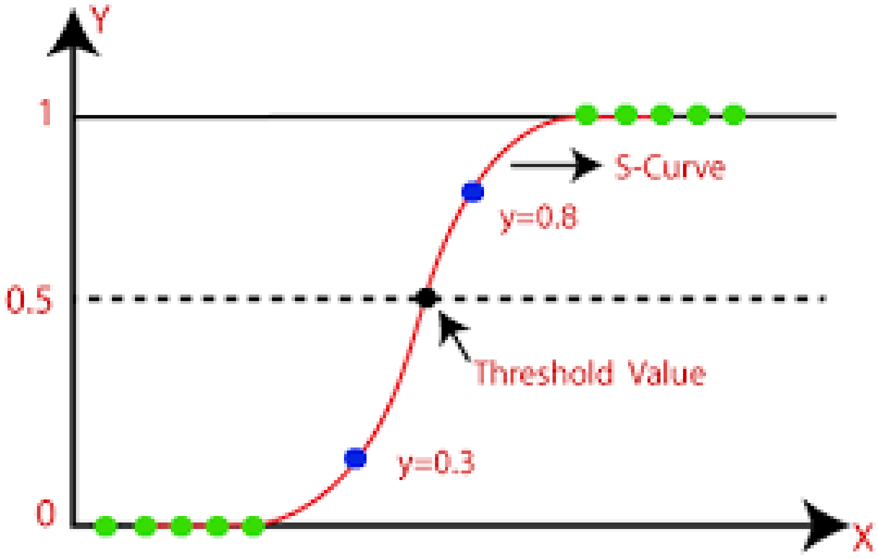


Fig 3. Logistic Regression Chart

3.8.2 Decision Tree

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
- The decisions or the test are performed on the basis of features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.
- A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

Accuracy Score of Credit Card Fraud Analysis using Decision Tree is : 89.84771573604061

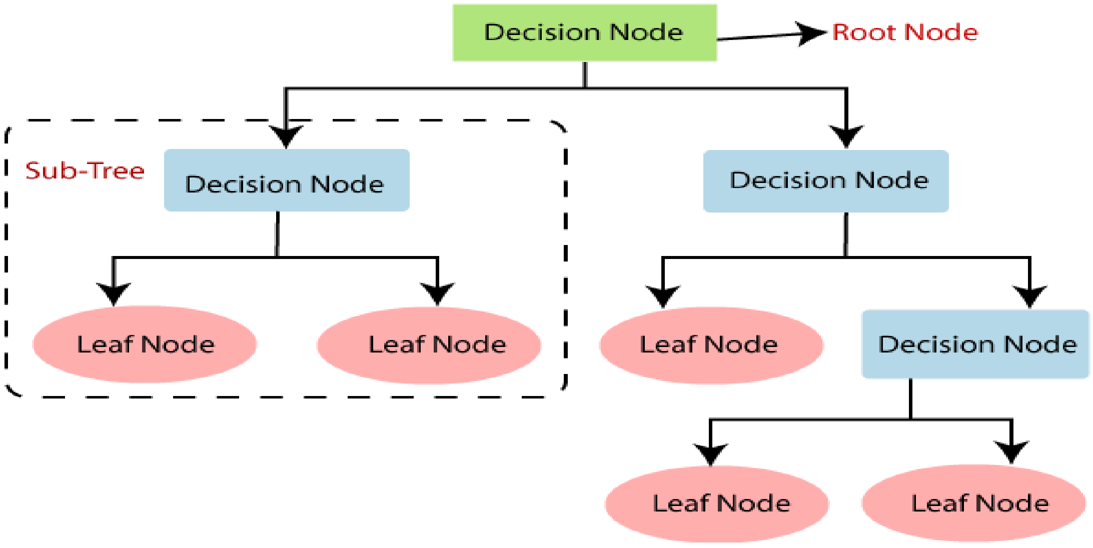


Fig 4. Decision Tree chart

3.8.3 Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning**,** which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

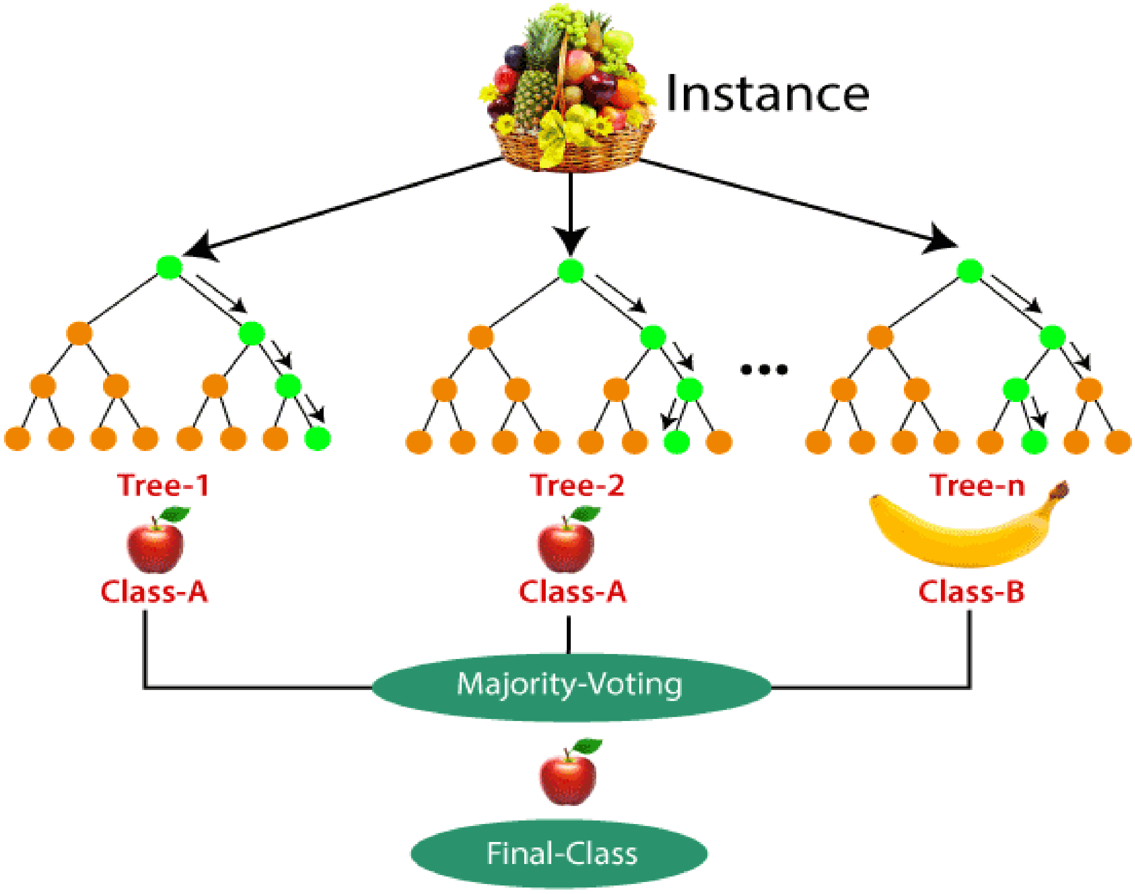


Fig 5. Random Forest Chart

Random forest has nearly the same hyperparameters as a decision tree or a bagging classifier. Fortunately, there's no need to combine a decision tree with a bagging classifier because you can easily use the classifier-class of random forest. With random forest, you can also deal with regression tasks by using the algorithm's regressor.

Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

Therefore, in random forest, only a random subset of the features is taken into consideration by the algorithm for splitting a node. You can even make trees more random by additionally using random thresholds for each feature rather than searching for the best possible thresholds (like a normal decision tree does).

Accuracy Score of Credit Card Fraud Analysis using Random Forest is : 92.38578680203046

3.8.4 K-Nearest Neighbor

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

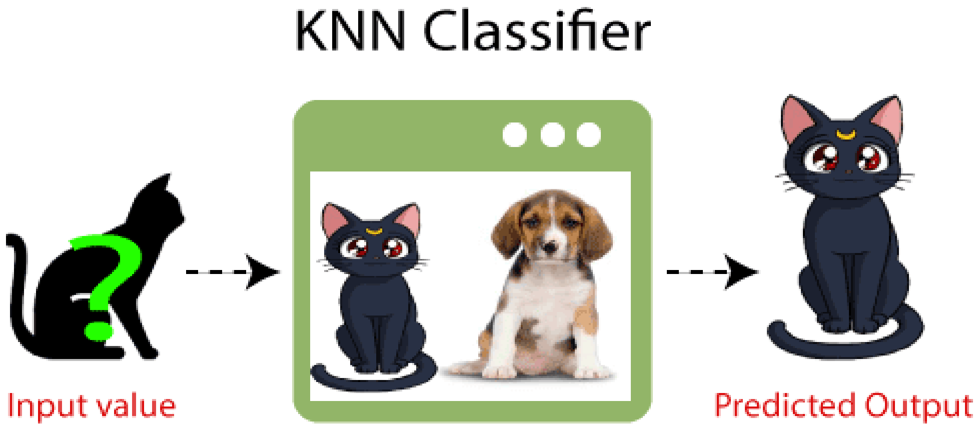


Fig 6. K-Nearest Neighbor Chart

- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.
- It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- K-NN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.
- **Ex:** Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the K-NN algorithm, as it works on a similarity measure. Our K-NN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

Accuracy Score for Credit Card Fraud Analysis using K-NN is : 62.94416243654822

3.9 Block Diagram & Methodology

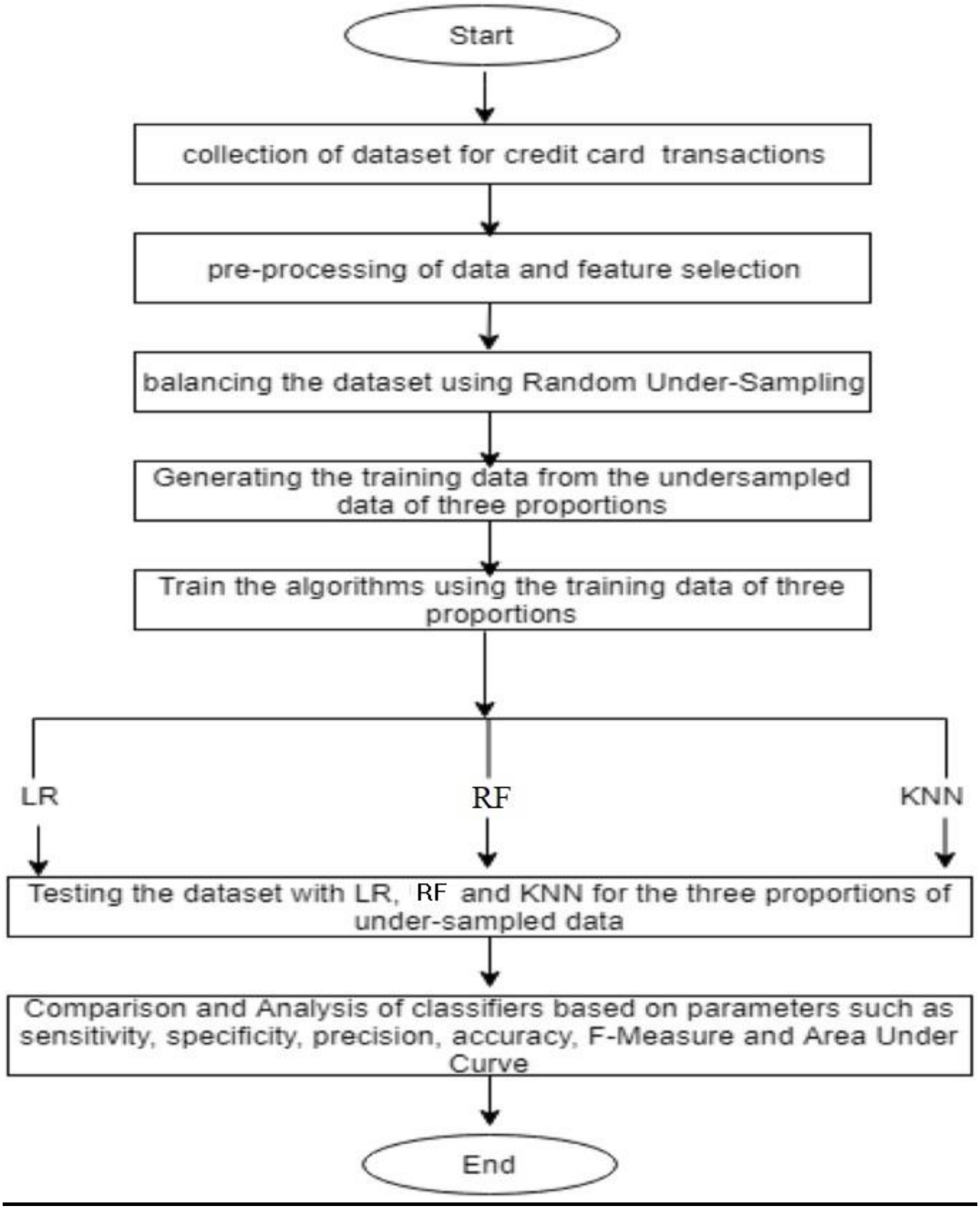

Start
collection of dataset for credit card transactions
pre-processing of data and feature selection
balancing the dataset using Random Under-Sampling
Generating the training data from the undersampled data of three proportions
Train the algorithms using the training data of three proportions
LR
RF
KNN
Testing the dataset with LR, RF and KNN for the three proportions of under-sampled data
Comparison and Analysis of classifiers based on parameters such as sensitivity, specificity, precision, accuracy, F-Measure and Area Under Curve
End

Fig 7. Block Diagram of Credit Card Fraud Detection using Different Algorithms

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but

instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1 values.

- Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.
- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).
- The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.
- Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
- Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification. The below image is showing the logistic function

3.9.1 Logistic Function (Sigmoid Function):

- The sigmoid function is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of 0 and 1.
- The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

Assumptions for Logistic Regression:

- The dependent variable must be categorical in nature.
- The independent variable should not have multi-collinearity.

3.9.2 Type of Logistic Regression:

On the basis of the categories, Logistic Regression can be classified into three types:

- **Binomial:** In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.

- **Multinomial:** In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"
- **Ordinal:** In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

Steps in Logistic Regression:

To implement the Logistic Regression using Python, we will use the same steps as we have done in previous topics of Regression. Below are the steps:

- Data Pre-processing step
- Fitting Logistic Regression to the Training set
- Predicting the test result
- Test accuracy of the result(Creation of Confusion matrix)
- Visualizing the test set result.

CHAPTER 4

HARDWARE AND SOFTWARE REQUIREMENTS

4.1 Hardware Requirements:

- RAM: 4GB and Higher
- Processor: Intel i3 and above
- Hard Disk: 500GB: Minimum

4.2 Software Requirements:

- OS: Windows or Linux
- Python IDE : python 2.7.x and above
- PyCharm IDE Required, Jupiter notebook
- Setup tools and pip to be installed for 3.6 and above
- Language : Python Scripting

4.2.1 Python Features:

Python and its Salient Features. Python is a dynamic, high-level, free open source, and interpreted programming language. It supports object-oriented programming as well as procedural-oriented programming. In Python, we don't need to declare the type of variable because it is a dynamically typed language.

1. **Easy to code.**

   Python is a high-level programming language. Python is very easy to learn language as compared to other languages like C, C#, JavaScript, Java, etc. It is very easy to code in the python language and anybody can learn python basics in a few hours or days. It is also a developer-friendly language.

2. **Free and Open Source.**

   Python language is freely available on the official website Since it is open-source, this means that source code is also available to the public. So, you can download it, use it as well as share it.

3. **Object-Oriented Language**.

   One of the key features of python is Object-Oriented programming. Python supports object-oriented language and concepts of classes, objects encapsulation, etc.

4. **GUI Programming Support.**

   Graphical User interfaces can be made using a module such as PyQt5, PyQt4, wx Python, or Tk in python. PyQt5 is the most popular option for creating graphical apps with Python.

5. **High-Level Language.**

   Python is a high-level language. When we write programs in python, we do not need to remember the system architecture, nor do we need to manage the memory.

6. **Extensible feature.**

   Python is an Extensible language. We can write some Python code into C or C++ language and also, we can compile that code in C/C++ language.

7. **Python is a Portable language.**

   Python language is also a portable language. For example, if we have python code for windows and if we want to run this code on other platforms such as Linux, Unix, and Mac then we do not need to change it, we can run this code on any platform.

8. **Python is an Integrated language.**

   Python is also an Integrated language because we can easily integrated python with other languages like C, C++, etc.

9. **Interpreted Language.**

   Python is an Interpreted Language because Python code is executed line by line at a time. like other languages C, C++, Java, etc. there is no need to compile python code this makes it easier to debug our code. The source code of python is converted into an immediate form called bytecode.

10. **Large Standard Library.**

    Python has a large standard library that provides a rich set of modules and functions so you do not have to write your code for every single thing. There are many libraries present in python for such as regular expressions, unit-testing, web browsers, etc.

11. **Dynamically Typed Language.**

    Python is a dynamically typed language. That means the type (for example- int, double, long, etc.) for a variable is decided at run time not in advance because of this feature we don't need to specify the type of variable.

4.2.2 Python Libraries

The Libraries used in this model are :

1. **Scikit-learn (Sklearn)**

   Sklearn is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

2. Pandas

Pandas is an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc. In this tutorial, we will learn the various features of Python Pandas and how to use them in practice.

3. Matplotlib

Matplotlib is a popular Python library that can be used to create data visualizations quite easily. It is probably the single most used Python package for 2D-graphics along with limited support for 3D-graphics. It provides both, a very quick way to visualize data from Python and publication-quality figures in many formats. Also, It was designed from the beginning to serve two purposes:

- Allow for interactive, cross-platform control of figures and plots.
- Make it easy to produce static vector graphics files without the need for any GUIs.

CHAPTER 5

RESULT DISCUSSION

Credit Card Fraud Detection Using Logistic Regression Algorithm**,** In this project we are using python Logistic Regression algorithm to detect fraud transaction from credit card dataset, we downloaded this dataset from "Kaggle's" web site from below URL. Dataset URL: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

To provide privacy to users transaction data Kaggle's peoples have converted transaction data to numerical format using PCA Algorithm. Below are some example from dataset **"Time","V1","V2","V3","V4","V5","V6","V7","V8","V9","V10","V11","V12","V13","V14","V15","V16","V17","V18","V19","V20","V21","V22","V23","V24","V25","V26","V27","V28","Amount", "Class"**

0, -1.3598071336738, -0.0727811733098497, 2.53634673796914, 1.37815522427443, -0.338320769942518, 0.462387777762292, 0.239598554061257, 0.0986979012610507, 0.363786969611213, 0.0907941719789316, -0.551599533260813, -0.617800855762348, -0.991389847235408, -0.311169353699879, 1.46817697209427, -0.470400525259478, 0.207971241929242, 0.0257905801985591, 0.403992960255733, 0.251412098239705,-0.018306777944153, 0.277837575558899, -0.110473910188767, 0.0669280749146731, 0.128539358273528, -0.189114843888824, 0.133558376740387, -0.0210530534538215, 149.62

0,1.19185711131486,0.26615071205963,0.16648011335321,0.448154078460911,0.0600176492822243,-0.0823608088155687,-0.0788029833323113,0.0851016549148104,-0.255425128109186,-0.166974414004614,1.61272666105479,1.06523531137287,0.48909501589608,-0.143772296441519,0.635558093258208,0.463917041022171,-0.114804663102346,-0.183361270123994,-0.145783041325259,-0.0690831352230203,-0.225775248033138,-0.638671952771851,0.101288021253234,-0.339846475529127,0.167170404418143,0.125894532368176,-0.00898309914322813,0.0147241691924927,2.69

406,-2.3122265423263,1.95199201064158,-1.60985073229769,3.9979055875468,-0.522187864667764,-1.42654531920595,-2.53738730624579,1.39165724829804,-2.77008927719433,-2.77227214465915,3.20203320709635,-2.89990738849473,-0.595221881324605,-4.28925378244217,0.389724120274487,-1.14074717980657,-2.83005567450437,-0.0168224681808257,0.416955705037907,0.126910559061474,0.517232370861764,-

```
0.0350493686052974,0.465211076182388,0.320198198514526,0.0445191674731724,0.177839798284401,
0.261145002567677,-0.143275874698919,0

406,-2.3122265423263,1.95199201064158,-1.60985073229769,3.9979055875468,-0.522187864667764,-
1.42654531920595,-2.53738730624579,1.39165724829804,-2.77008927719433,-
2.77227214465915,3.20203320709635,-2.89990738849473,-0.595221881324605,-
4.28925378244217,0.389724120274487,-1.14074717980657,-2.83005567450437,-
0.0168224681808257,0.416955705037907,0.126910559061474,0.517232370861764,-
0.0350493686052974,0.465211076182388,0.320198198514526,0.0445191674731724,0.177839798284401,
0.261145002567677,-0.143275874698919,0


391,0.829932199150855,0.430250426288754,1.3037076726496,1.04055937115489,0.822209790945024,0.
844830135248345,0.589865274641566,0.0457907435900726,-
0.148644570189748,0.310522866846856,0.178932310526362,-0.314635572366816,-
1.51005928782656,0.18233120793819,-0.157713526599735,-
0.783954373578739,0.0373907611680013,0.242676948994549,1.44209294812507,-0.0588157812103496,-
0.0751789705110257,0.0960381661730585,-0.139414587240587,-
0.859739201249434,0.121235410228106,-0.204961889048451,-0.292247194199781,-
0.0589955070930744,27.7


392,-2.15349086514568,0.104020929476975,1.45953150222605,-1.61328386562366,-
1.07136061911647,0.192161000465908,-0.332747787806063,0.462717462316893,-
0.209569473605167,0.237995865802194,-1.6369911449172,-0.145006279441582,1.1946279388561,-
1.27675066350071,-1.22884605434653,1.18370927043986,0.201087890011779,-
1.32382362567056,0.676053146968067,0.227422001563131,-0.101936789226138,0.0967345142072667,-
0.490325274982157,-0.392662347164788,0.608577595433795,-0.338381669720943,-0.352131899441536,-
0.422725981644548,82.29


393,-2.59549971893285,1.31159181560757,1.32308130712122,-1.1278489024485,0.307467369283372,-
0.518201824511376,-2.09881663231913,-
4.38246895752246,1.72453455583367,0.198679921779162,0.381591794293992,0.780889731957347,-
0.247810881747048,0.0469023001974929,2.50139470225836,-1.43212684226367,0.732979440811641,-
0.183239271138246,1.07563087678484,-1.14867430422767,3.98038383296976,-0.835844602449599,-
1.78632894189546,0.485594402321421,-0.535158060477884,-0.684143186059059,-
```

0.42254203234575,0.0470780379929637,64.04

6986,-4.39797444171999,1.35836702839758,-2.5928442182573,2.67978696694832,-1.12813094208956,-1.70653638774951,-3.49619729302467,-0.248777743025673,-0.24776789948008,-4.80163740602813,4.89584422347523,-10.9128193194019,0.184371685834387,-6.77109672468083,-0.00732618257771211,-7.35808322132346,-12.5984185405511,-5.13154862842983,0.308333945758691,-0.17160787864796,0.573574068424352,0.176967718048195,-0.436206883597401,-0.0535018648884285,0.252405261951833,-0.657487754764504,-0.827135714578603,0.849573379985768,59

6982,1.08920512319588,-0.218264395186058,1.10478804117043,1.17478002950539,-0.86829344684362,-0.0256698527090968,-0.562192423083571,0.0764778624647165,2.35693940010332,-0.725093737457287,0.348622212582567,-2.13013025643224,0.670256846166844,1.15957256033765,-1.31493320080899,-0.858892373160103,1.43384187762072,-0.820262053006578,-0.326410299691067,-0.210565238613142,-0.240386165442438,

0.1695193041708,0.00631852875949629,0.389350972524518,0.3373566008662,0.498937247878272,-0.0244941197637929,0.00931744739955896,31.81

6982,0.965085036323121,-0.539907382654771,0.426387593426602,0.705498809294599,-0.587330942596624,-0.0475559980038503,-0.357081760772518,-0.0873014326700456,1.92071747849171,-0.511201685299222,-0.443104042479314,-3.05397443668305,1.29121473008275,1.54426737267137,0.724752716974505,0.834499340177594,-0.100061175828153,0.857971332954157,-0.320175933764894,0.215026930061365,0.070183714661053,0.0896547813743098,-0.338481158757214,-0.583627938037564,0.399460678795733,0.621477774448982,-0.0781027401317497,0.0325123056509956,180.16

In the page 39 bold names are the column names of this dataset and others decimal values are the content of dataset and in above 3 rows last column contains class label where 0 means transaction values are normal and 1 means contains fraud values.

Using above “CreditCardFraud.csv” file we will train Logistic Regression algorithm and then we will upload test data file and this test data will be applied on Logistic Regression train model to predict whether test data contains normal or fraud transaction signatures. When we upload test data then it will contains only transaction data no class label will be there application will predict and give the result.

In above screen in test data file there are no 0 or 1 values, application will predict from this test data using random forest and give the result.

Screenshots:

- To run project double click on "**run**" button on any idle python editor

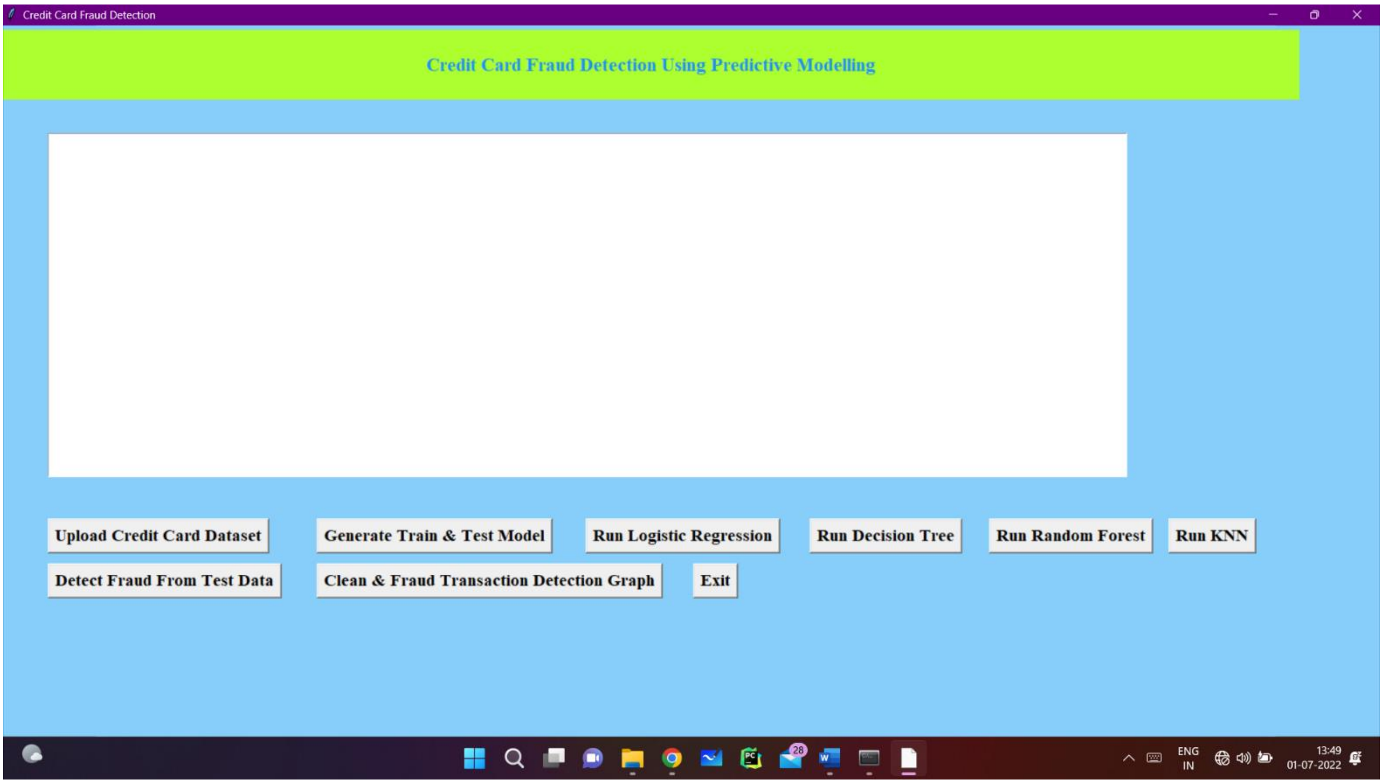
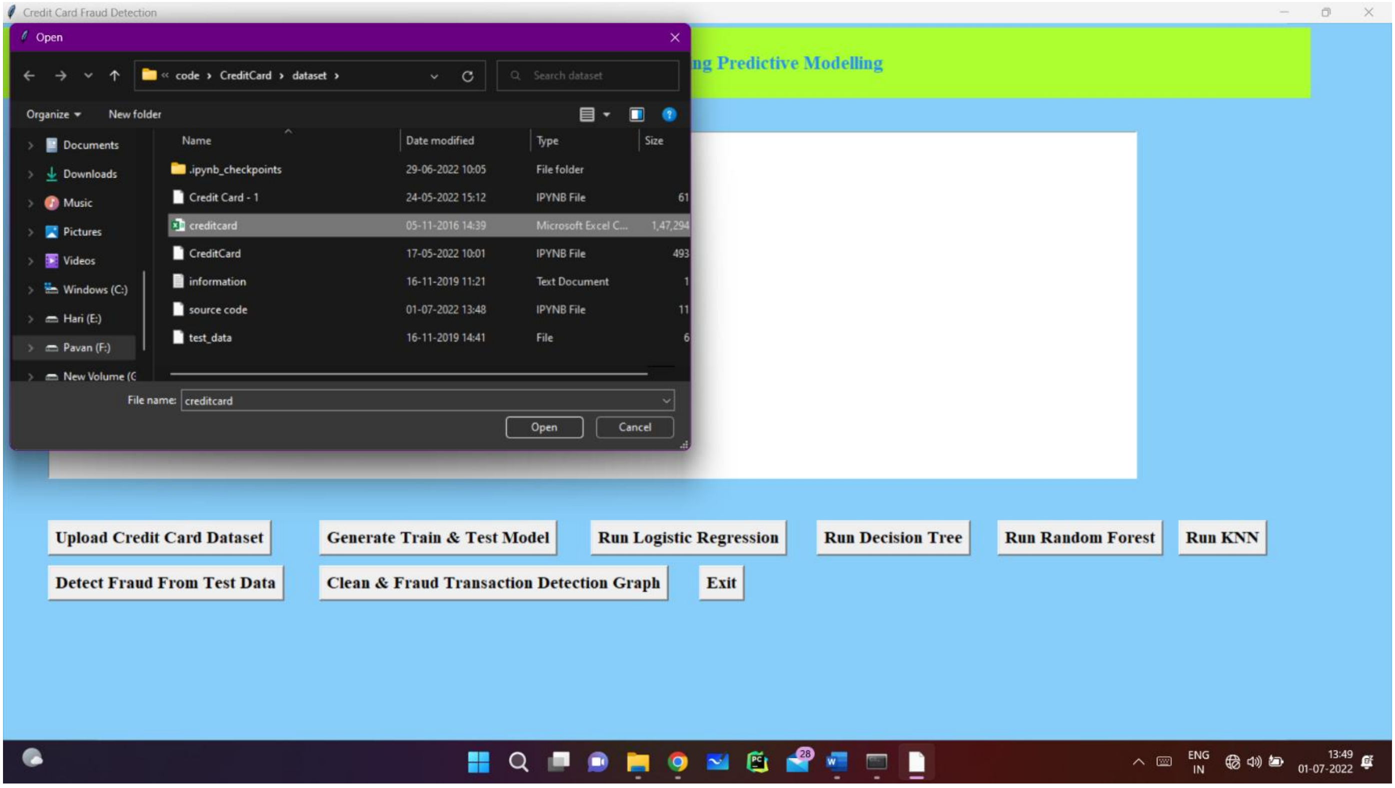

Credit Card Fraud Detection
Credit Card Fraud Detection Using Predictive Modelling
Upload Credit Card Dataset
Generate Train & Test Model
Run Logistic Regression
Run Decision Tree
Run Random Forest
Run KNN
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
ENG
IN
13:49
01-07-2022

Fig 8.1 GUI for running program

- In above screen click on "**Upload Credit Card Dataset**" button to upload dataset

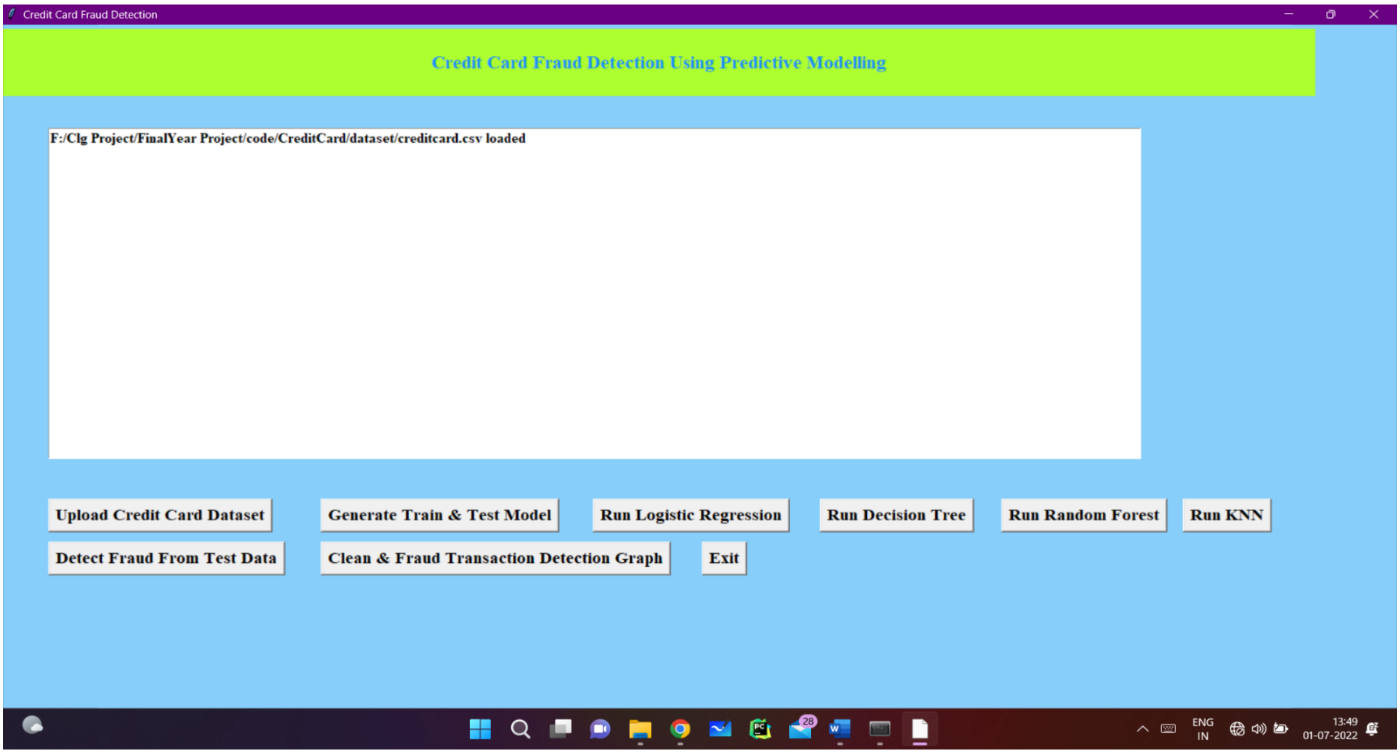


Name	Date modified	Type	Size
.ipynb_checkpoints	29-06-2022 10:05	File folder	
Credit Card - 1	24-05-2022 15:12	IPYNB File	61
creditcard	05-11-2016 14:39	Microsoft Excel C...	1,47,294
CreditCard	17-05-2022 10:01	IPYNB File	493
information	16-11-2019 11:21	Text Document	1
source code	01-07-2022 13:48	IPYNB File	11
test_data	16-11-2019 14:41	File	6

File name: creditcard

Fig 8.2 GUI for uploading file

- After uploading dataset will get below screen

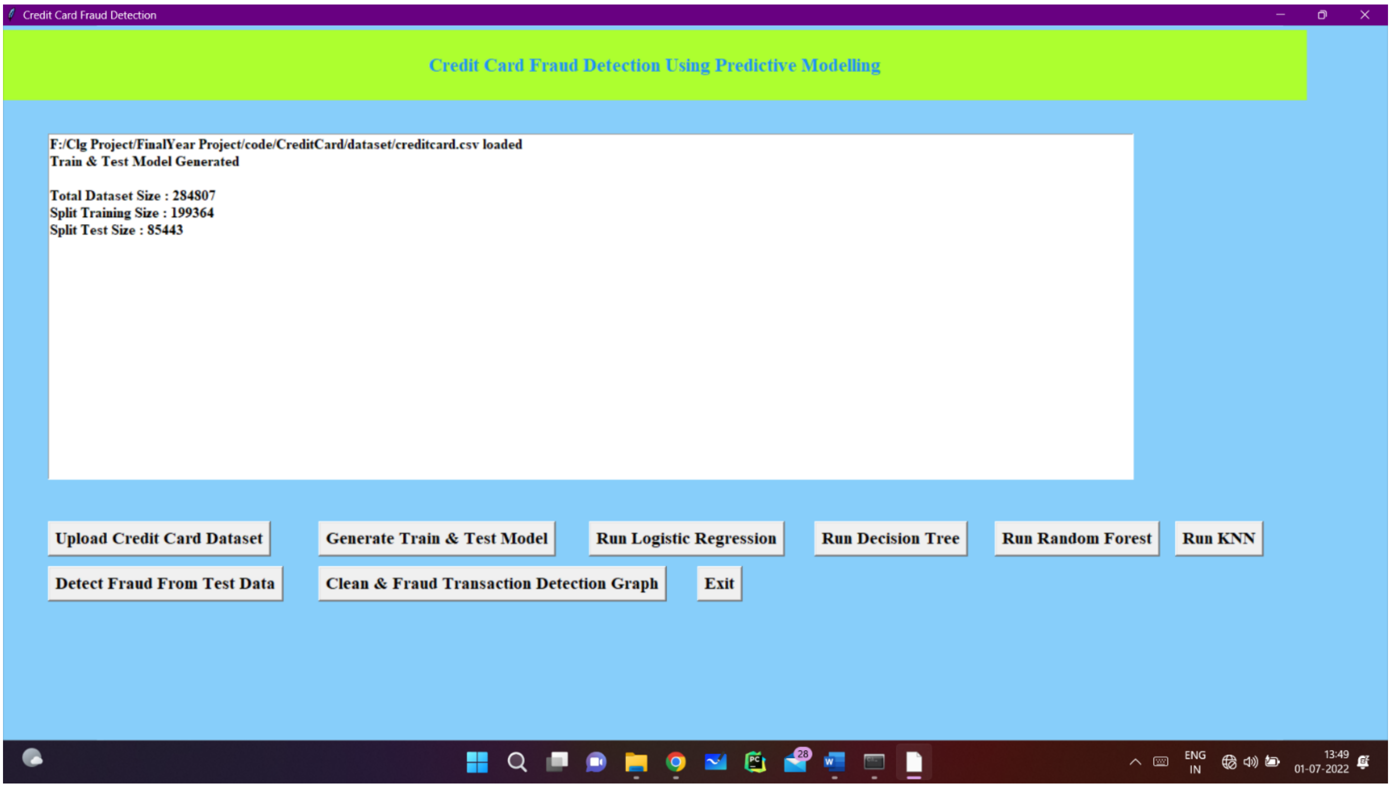


Credit Card Fraud Detection
Credit Card Fraud Detection Using Predictive Modelling
Upload Credit Card Dataset
Generate Train & Test Model
Run Logistic Regression
Run Decision Tree
Run Random Forest
Run KNN
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
ENG
IN
13:49
01-07-2022

F:/Clg Project/FinalYear Project/code/CreditCard/dataset/creditcard.csv loaded

Fig 8.3 GUI indicates the dataset loaded

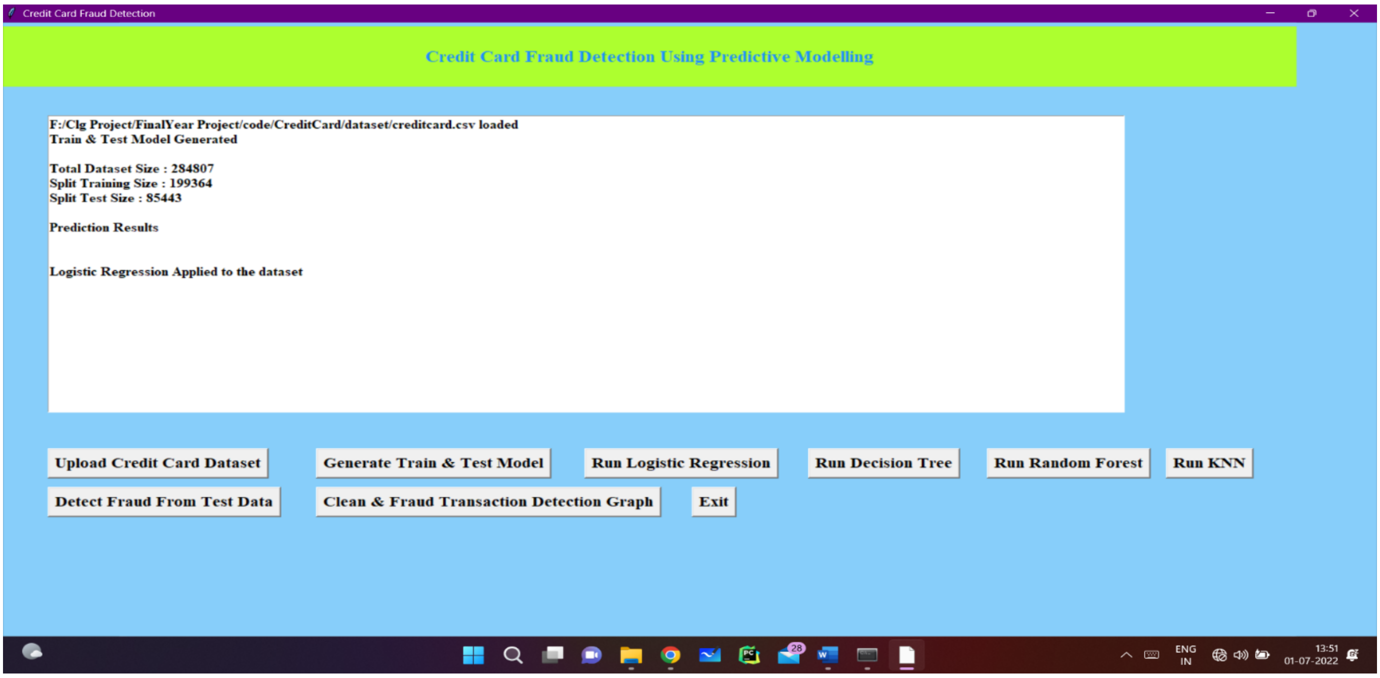
- Now click on “**Generate Train & Test Model**” to generate training model for Logistic Regression


Credit Card Fraud Detection
Credit Card Fraud Detection Using Predictive Modelling
Upload Credit Card Dataset
Generate Train & Test Model
Run Logistic Regression
Run Decision Tree
Run Random Forest
Run KNN
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
ENG
IN
13:49
01-07-2022

```
F:/Clg Project/FinalYear Project/code/CreditCard/dataset/creditcard.csv loaded
Train & Test Model Generated

Total Dataset Size : 284807
Split Training Size : 199364
Split Test Size : 85443
```

Fig 8.4 GUI splits the data for training & testing

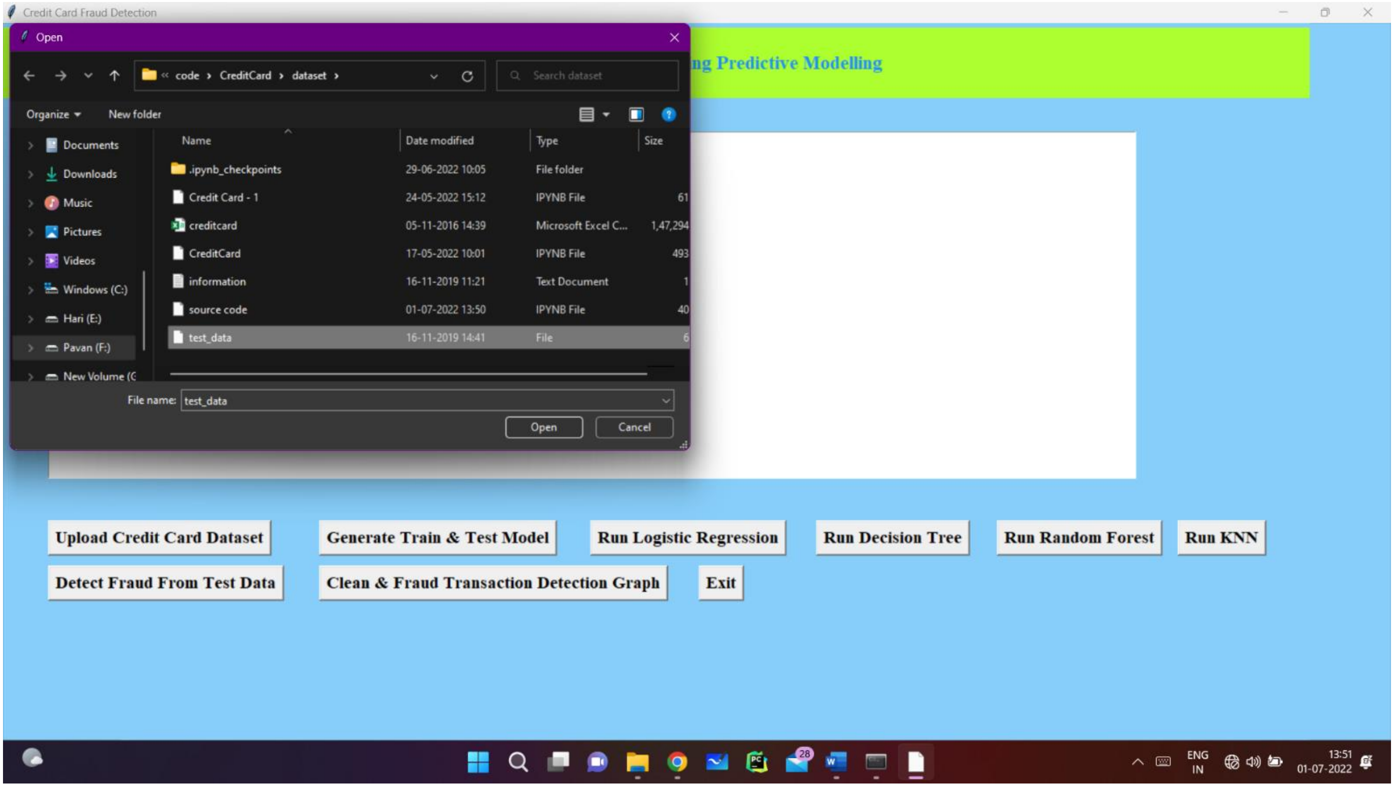
- In above screen after generating model we can see total records available in dataset and then application using how many records for training and how many for testing. Now click on "Run Logistic Regression" button to generate Logistic Regression model on train and test data.


Credit Card Fraud Detection
Credit Card Fraud Detection Using Predictive Modelling
Upload Credit Card Dataset
Generate Train & Test Model
Run Logistic Regression
Run Decision Tree
Run Random Forest
Run KNN
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
ENG
IN
13:51
01-07-2022

```
F:/Clg Project/FinalYear Project/code/CreditCard/dataset/creditcard.csv loaded
Train & Test Model Generated

Total Dataset Size : 284807
Split Training Size : 199364
Split Test Size : 85443

Prediction Results


Logistic Regression Applied to the dataset
```

Fig 8.5 GUI to select the Logistic Regression

- In above screen we can see Logistic Regression is generated while building model on train and test data. Now click on "**Detect Fraud from Test Data**" button to upload test data and to predict whether test data contains normal or fraud transaction

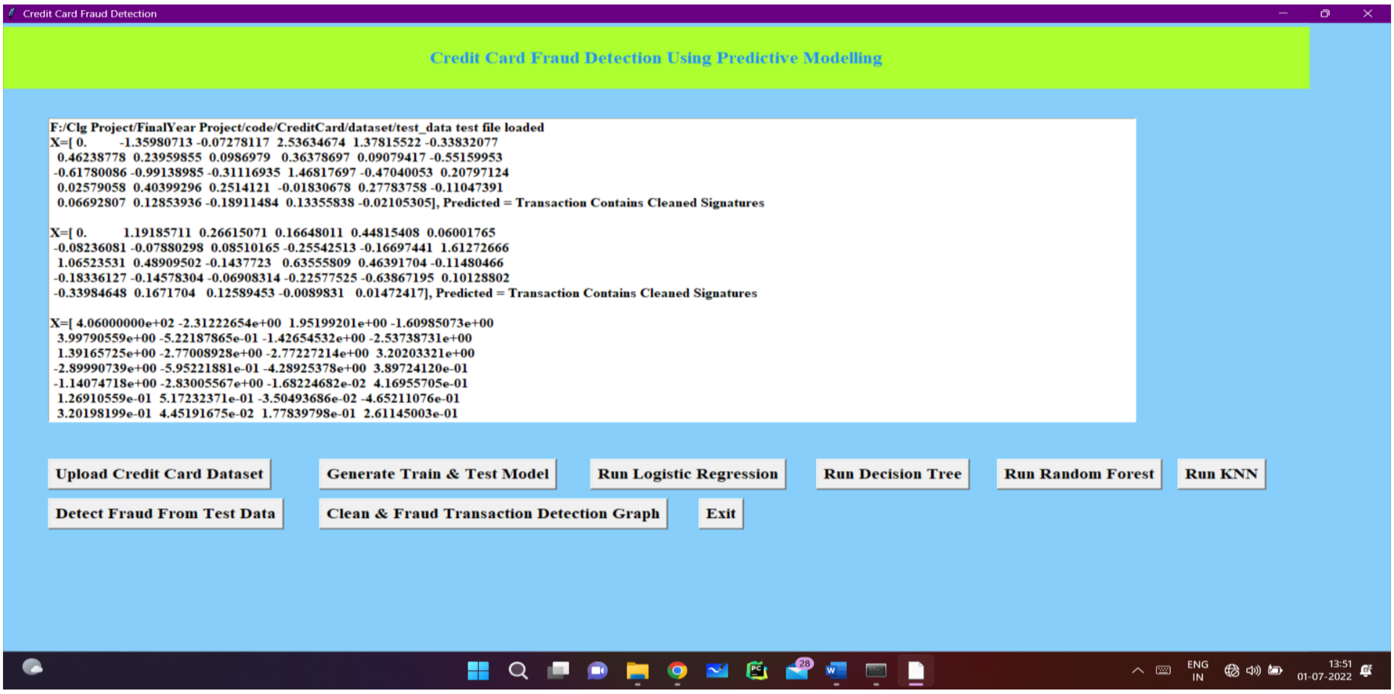


Name	Date modified	Type	Size
.ipynb_checkpoints	29-06-2022 10:05	File folder	
Credit Card - 1	24-05-2022 15:12	IPYNB File	61
creditcard	05-11-2016 14:39	Microsoft Excel C...	1,47,294
CreditCard	17-05-2022 10:01	IPYNB File	493
information	16-11-2019 11:21	Text Document	1
source code	01-07-2022 13:50	IPYNB File	40
test_data	16-11-2019 14:41	File	6

File name: test_data

Fig 8.6 GUI indicates Detect Fraud from test data

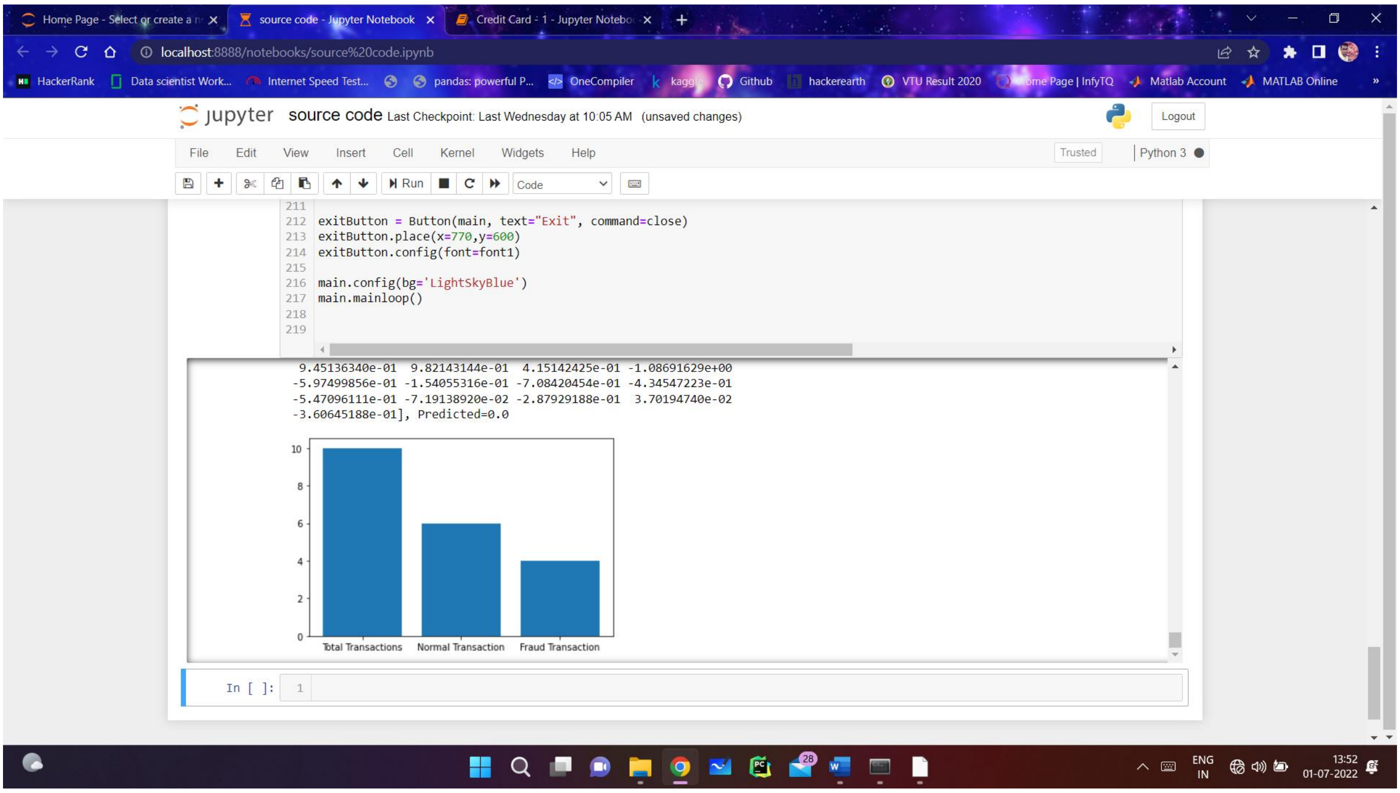
- In above screen I am uploading test dataset and after uploading test data will get below prediction details


Credit Card Fraud Detection
Credit Card Fraud Detection Using Predictive Modelling
Upload Credit Card Dataset
Generate Train & Test Model
Run Logistic Regression
Run Decision Tree
Run Random Forest
Run KNN
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
ENG
IN
13:51
01-07-2022

```
F:/Clg Project/FinalYear Project/code/CreditCard/dataset/test_data test file loaded
X=[ 0.         -1.35980713 -0.07278117  2.53634674  1.37815522 -0.33832077
  0.46238778  0.23959855  0.0986979   0.36378697  0.09079417 -0.55159953
 -0.61780086 -0.99138985 -0.31116935  1.46817697 -0.47040053  0.20797124
  0.02579058  0.40399296  0.2514121  -0.01830678  0.27783758 -0.11047391
  0.06692807  0.12853936 -0.18911484  0.13355838 -0.02105305], Predicted = Transaction Contains Cleaned Signatures

X=[ 0.          1.19185711  0.26615071  0.16648011  0.44815408  0.06001765
 -0.08236081 -0.07880298  0.08510165 -0.25542513 -0.16697441  1.61272666
  1.06523531  0.48909502 -0.1437723   0.63555809  0.46391704 -0.11480466
 -0.18336127 -0.14578304 -0.06908314 -0.22577525 -0.63867195  0.10128802
 -0.33984648  0.1671704   0.12589453 -0.0089831   0.01472417], Predicted = Transaction Contains Cleaned Signatures

X=[ 4.06000000e+02 -2.31222654e+00  1.95199201e+00 -1.60985073e+00
  3.99790559e+00 -5.22187865e-01 -1.42654532e+00 -2.53738731e+00
  1.39165725e+00 -2.77008928e+00 -2.77227214e+00  3.20203321e+00
 -2.89990739e+00 -5.95221881e-01 -4.28925378e+00  3.89724120e-01
 -1.14074718e+00 -2.83005567e+00 -1.68224682e-02  4.16955705e-01
  1.26910559e-01  5.17232371e-01 -3.50493686e-02 -4.65211076e-01
  3.20198199e-01  4.45191675e-02  1.77839798e-01  2.61145003e-01
```

Fig 8.7 GUI indicating normal and fraud transaction

- In above screen beside each test data application will display output as whether transaction contains cleaned or fraud signatures. Now click on “**Clean & Fraud Transaction Detection Graph**” button to see total test transaction with clean and fraud signature in graphical format. See below screen



```

exitButton = Button(main, text="Exit", command=close)
exitButton.place(x=770,y=600)
exitButton.config(font=font1)

main.config(bg='LightSkyBlue')
main.mainloop()


```

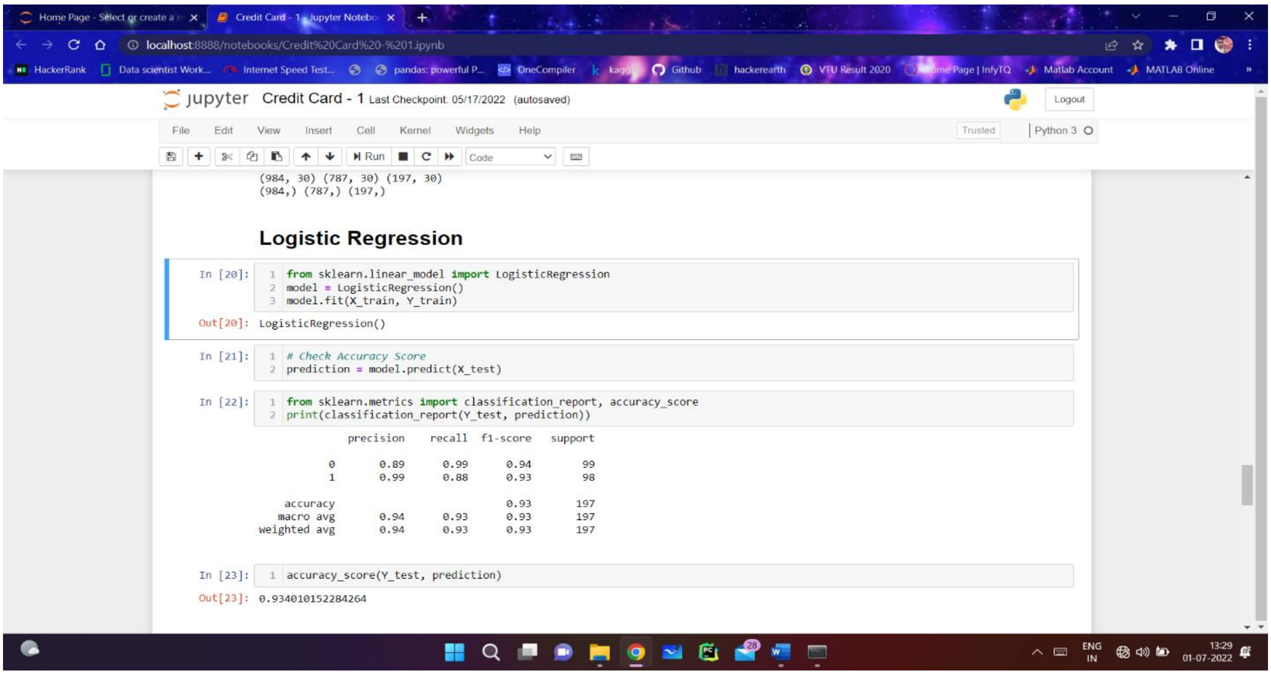
```
  9.45136340e-01  9.82143144e-01  4.15142425e-01 -1.08691629e+00
 -5.97499856e-01 -1.54055316e-01 -7.08420454e-01 -4.34547223e-01
 -5.47096111e-01 -7.19138920e-02 -2.87929188e-01  3.70194740e-02
 -3.60645188e-01], Predicted=0.0
```

Fig 8.8 GUI for Graphical representation

- In above graph we can see total test data and number of normal and fraud transaction detected. In above graph x-axis represents type and y-axis represents count of clean and fraud transaction.

Accuracy score of models:

- Logistic Regression



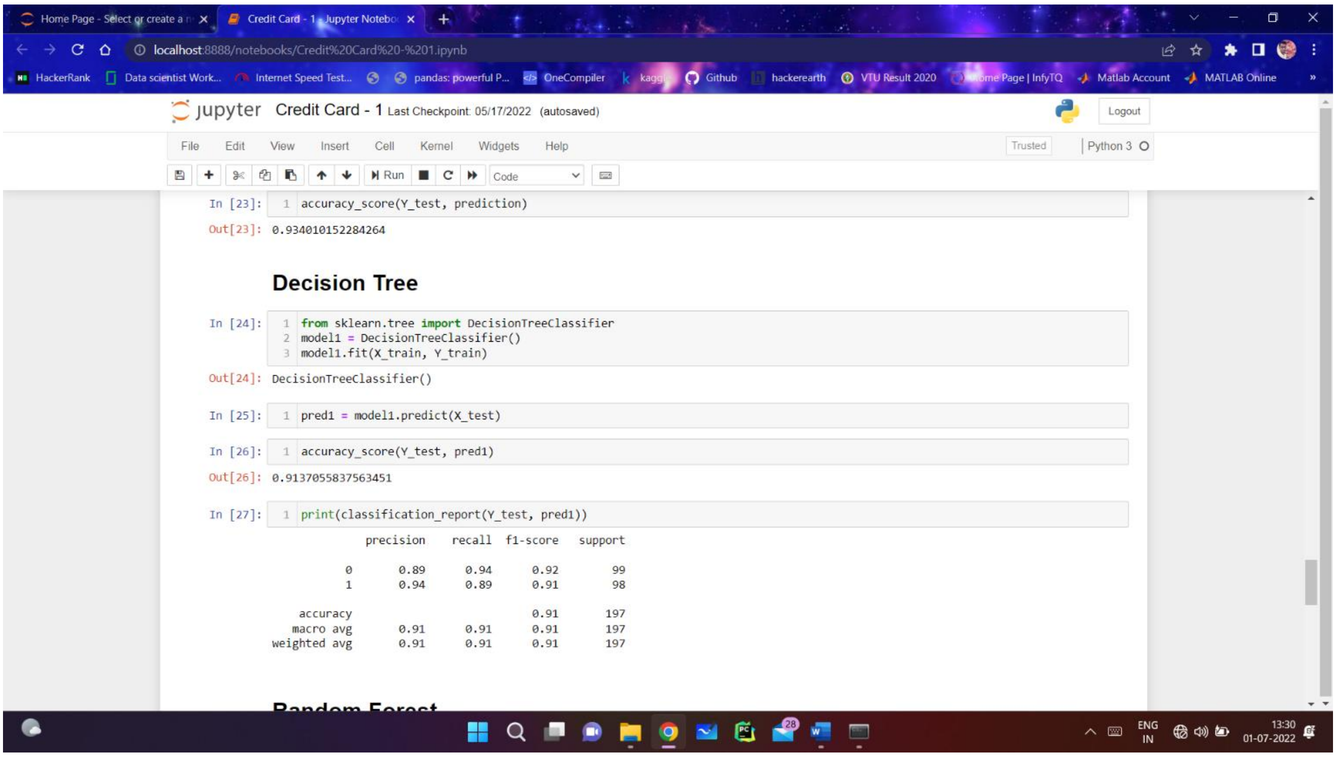
```
(984, 30) (787, 30) (197, 30)
(984,) (787,) (197,)
```

Logistic Regression

```
In [20]: from sklearn.linear_model import LogisticRegression
         model = LogisticRegression()
         model.fit(X_train, Y_train)

Out[20]: LogisticRegression()

In [21]: # Check Accuracy Score
         prediction = model.predict(X_test)

In [22]: from sklearn.metrics import classification_report, accuracy_score
         print(classification_report(Y_test, prediction))

              precision    recall  f1-score   support

           0       0.89      0.99      0.94        99
           1       0.99      0.88      0.93        98

    accuracy                           0.93       197
   macro avg       0.94      0.93      0.93       197
weighted avg       0.94      0.93      0.93       197


In [23]: accuracy_score(Y_test, prediction)

Out[23]: 0.934010152284264
```

Fig 8.9 Accuracy score of Logistic Regression

- Decision Tree


localhost:8888/notebooks/Credit%20Card%20-%201.ipynb
jupyter Credit Card - 1 Last Checkpoint: 05/17/2022 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code

```
In [23]: accuracy_score(Y_test, prediction)
Out[23]: 0.934010152284264
```

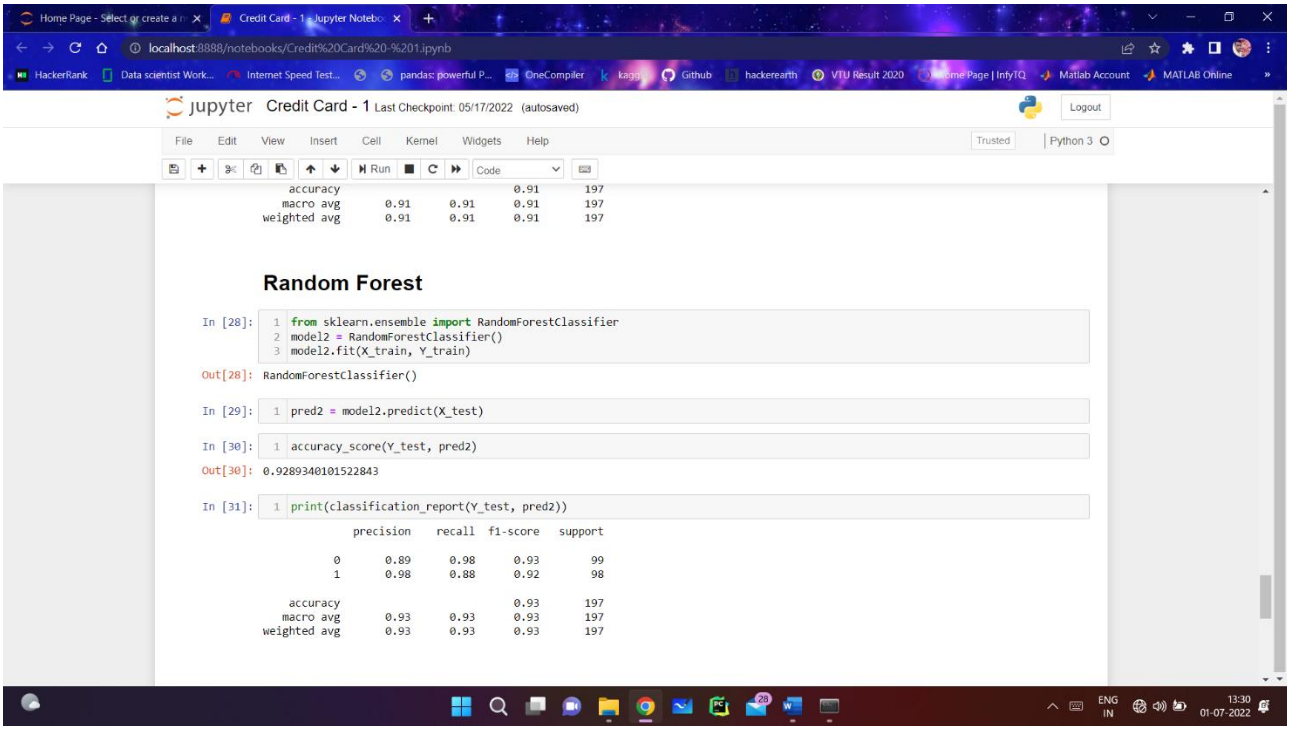
Decision Tree

```
In [24]: from sklearn.tree import DecisionTreeClassifier
         model1 = DecisionTreeClassifier()
         model1.fit(X_train, Y_train)
Out[24]: DecisionTreeClassifier()

In [25]: pred1 = model1.predict(X_test)

In [26]: accuracy_score(Y_test, pred1)
Out[26]: 0.9137055837563451

In [27]: print(classification_report(Y_test, pred1))
              precision    recall  f1-score   support

           0       0.89      0.94      0.92        99
           1       0.94      0.89      0.91        98

    accuracy                           0.91       197
   macro avg       0.91      0.91      0.91       197
weighted avg       0.91      0.91      0.91       197
```

Random Forest

Fig 8.10 Accuracy score of Decision Tree

- Random Forest


localhost:8888/notebooks/Credit%20Card%20-%201.ipynb
jupyter Credit Card - 1 Last Checkpoint: 05/17/2022 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code

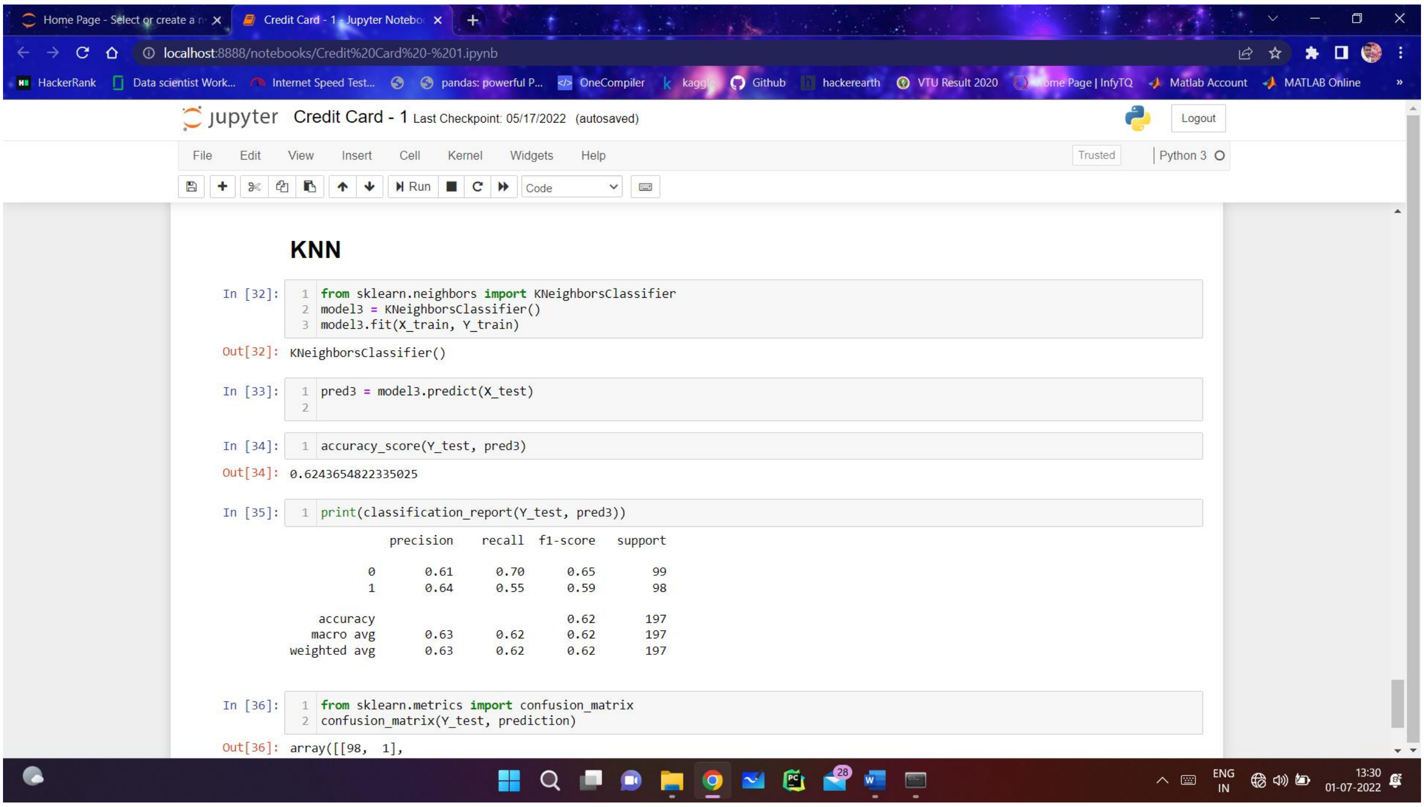
```
    accuracy                           0.91       197
   macro avg       0.91      0.91      0.91       197
weighted avg       0.91      0.91      0.91       197
```

Random Forest

```
In [28]: from sklearn.ensemble import RandomForestClassifier
         model2 = RandomForestClassifier()
         model2.fit(X_train, Y_train)
Out[28]: RandomForestClassifier()

In [29]: pred2 = model2.predict(X_test)

In [30]: accuracy_score(Y_test, pred2)
Out[30]: 0.9289340101522843

In [31]: print(classification_report(Y_test, pred2))
              precision    recall  f1-score   support

           0       0.89      0.98      0.93        99
           1       0.98      0.88      0.92        98

    accuracy                           0.93       197
   macro avg       0.93      0.93      0.93       197
weighted avg       0.93      0.93      0.93       197
```

Fig 8.11 Accuracy score of Random Forest

- K-Nearest Neighbor



KNN

```
In [32]: from sklearn.neighbors import KNeighborsClassifier
         model3 = KNeighborsClassifier()
         model3.fit(X_train, Y_train)
Out[32]: KNeighborsClassifier()

In [33]: pred3 = model3.predict(X_test)

In [34]: accuracy_score(Y_test, pred3)
Out[34]: 0.6243654822335025

In [35]: print(classification_report(Y_test, pred3))
                      precision    recall  f1-score   support

                   0       0.61      0.70      0.65        99
                   1       0.64      0.55      0.59        98

            accuracy                           0.62       197
           macro avg       0.63      0.62      0.62       197
        weighted avg       0.63      0.62      0.62       197

In [36]: from sklearn.metrics import confusion_matrix
         confusion_matrix(Y_test, prediction)
Out[36]: array([[98,  1],
```

Fig 8.12 Accuracy score of K-Nearest Neighbor

5.1 Observation:

S.No	Machine Learning Algorithm	Dataset	Accuracy score in %
1	Logistic Regression	Kaggle	93.40
2	Decision Tree		91.37
3	Random Forest		92.89
4	K-Nearest Neighbor		62.43

The best accuracy score obtained practically is 93.40% from Logistic Regression.

CHAPTER 6

APPLICATIONS AND ADVANTAGES

6.1 Applications:

From the moment the e-commerce payment systems came to existence, there have always been people who will find new ways to access someone's finances illegally. This has become a major problem in the modern era, as all transactions can easily be completed online by only entering your credit card information. Even in the 2010s, many American retail website users were the victims of online transaction fraud right before two-step verification was used for shopping online. Organizations, consumers, banks, and merchants are put at risk when a data breach leads to monetary theft and ultimately the loss of customers' loyalty along with the company's reputation.

Unauthorized card operations hit an astonishing amount of 16.7 million victims in 2017. Additionally, as reported by the Federal Trade Commission (FTC), the number of credit card fraud claims in 2017 was 40% higher than the previous year's number. There were around 13,000 reported cases in California and 8,000 in Florida, which are the largest states per capita for such type of crime. The amount of money at stake will exceed approximately $30 billion by 2020.

6.2 Advantages:

6.2.1 Keep customer credit history

Having a good credit history is often important in detecting loyal customers. This history is valuable not only for credit cards, but also for other financial services like loans, rental applications, or even some jobs. Lenders and issuers of credit mortgage companies, credit card companies, retail stores, and utility companies can review customer credit score and history to see how punctual and responsible customers are in paying back their debts.

6.2.2 Protection of Purchases

Credit cards may also offer customers, additional protection if the purchased Merchandise becomes lost, damaged, or stolen. Both the buyers credit card statement and company can confirm that the customer has bought if the original receipt is lost or stolen. In addition, some credit card companies provide insurance for large purchases.

CHAPTER 7

CONCLUSION

The detection of credit card fraud is a vital research field. This is because of the increasing number of fraud cases in financial institutions & online transactions. This issue opens the door for employing artificial intelligence to build systems that can detect fraud. Building an AI-based system to detect fraud requires a database to train the system (or classifier). The data in reality are dirty and have missing values, noisy data, and outliers. Such issues negatively affect the accuracy rate of the system. To overcome these problems, a logistic regression-based classifier is proposed. The data are first cleaned using two methods: the mean-based method and clustering-based method. Second, the classifier is trained based on the cross validation technique (folds=10), which ensures that the whole database is used as both the training data set and testing data set. Finally, the proposed classifier is evaluated based on the accuracy. The proposed logistic regression-based classifier is compared to well-known classifiers, which are the K-nearest neighbours , Decision Tree Classifier and Random Forest Classifier. The logistic regression-based classifier generates the best results (accuracy = 93.40%).

CHAPTER 8

REFERENCES

[1] P. Save, P. Tiwarekar, K. N., and N. Mahyavanshi, —A Novel Idea for Credit Card Fraud Detection using Decision Tree,‖ Int. J. Comput. Appl., vol. 161, no. 13, pp. 6–9, 2017, doi: 10.5120/ijca2017913413.

[2] Vimala Devi and K. S. Kavitha, —Fraud Detection in Credit Card Transactions by using Classification Algorithms,‖ Int. Conf. Curr. Trends Comput. Electr. Electron. Commun. CTCEEC 2017, pp. 125–131, 2018, doi: 10.1109/CTCEEC.2017.8455091.

[3] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, —Random forest for credit card fraud detection,‖ ICNSC 2018 - 15th IEEE Int. Conf. Networking, Sens. Control, pp. 1–6, 2018, doi: 10.1109/ICNSC.2018.8361343.

[4] R. R. Popat and J. Chaudhary, —A Survey on Credit Card Fraud Detection Using Machine Learning,‖ Proc. 2nd Int. Conf. Trends Electron. Informatics, ICOEI 2018, no. Icoei, pp. 1120–1125, 2018, doi: 10.1109/ICOEI.2018.8553963.

[5] V. Patil and U. Kumar Lilhore, —A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection,‖ Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol. © 2018 IJSRCSEIT, vol. 5, no. 10, pp. 320–325, 2018, doi: 10.13140/RG.2.2.22116.73608.

[6] S. Mittal and S. Tyagi, —Performance evaluation of machine learning algorithms for credit card fraud detection,‖ Proc. 9th Int. Conf. Cloud Comput. Data Sci. Eng. Conflu. 2019, pp. 320–324, 2019, doi: 10.1109/CONFLUENCE.2019.8776925.

[7] Yousefi, Niloofar, Marie Alaghband, and Ivan Garibay. "A Comprehensive Survey on Machine Learning Techniques and User Authentication Approaches for Credit Card Fraud Detection." arXiv preprint arXiv:1912.02629 (2019).

[8] X. Yu, X. Li, Y. Dong, and R. Zheng, —A Deep Neural Network Algorithm for Detecting Credit Card Fraud,‖ Proc. - 2020 Int. Conf. Big Data, Artif. Intell. Internet Things Eng. ICBAIE 2020, pp. 181–183, 2020, doi: 10.1109/ICBAIE49996.2020.00045.

[9] R. San Miguel Carrasco and M.-A. Sicilia-Urban, ―Evaluation of Deep Neural Networks for Reduction of Credit Card Fraud Alerts,‖ IEEE Access, vol. 8, pp. 186421–186432, 2020, doi: 10.1109/access.2020.3026222.

[10] S. H. Projects and W. Lovo, ―JMU Scholarly Commons Detecting credit card fraud : An analysis of fraud detection techniques,‖ 2020.

[11] Janbandhu, Ruchika, Shameedha Begum, and N. Ramasubramanian. "Credit Card Fraud Detection." Computing in Engineering and Technology. Springer, Singapore, 2020.

[12] Gianini, Gabriele, et al. "Managing a pool of rules for credit card fraud detection by a Game Theory based approach." Future Generation Computer Systems 102 (2020)

[13] Thabtah, Fadi, et al. "Data imbalance in classification: Experimental evaluation." Information Sciences 513 (2020)

[14] G. Kibria and M. Sevkli, ―Application of Deep Learning for Credit Card Approval : A Comparison with Application of Deep Learning for Credit Card Approval : A Comparison with Two Machine Learning Techniques,‖ no. January, pp. 0–5, 2021, doi: 10.18178/ijmlc.2021.11.4.1049.

[15]D. D. Borse, P. S. H. Patil, and S. Dhotre, ―Credit Card Fraud Detection Using Naïve Bayes and C4,‖ vol. 10, no. 1, pp. 423–429, 2021.