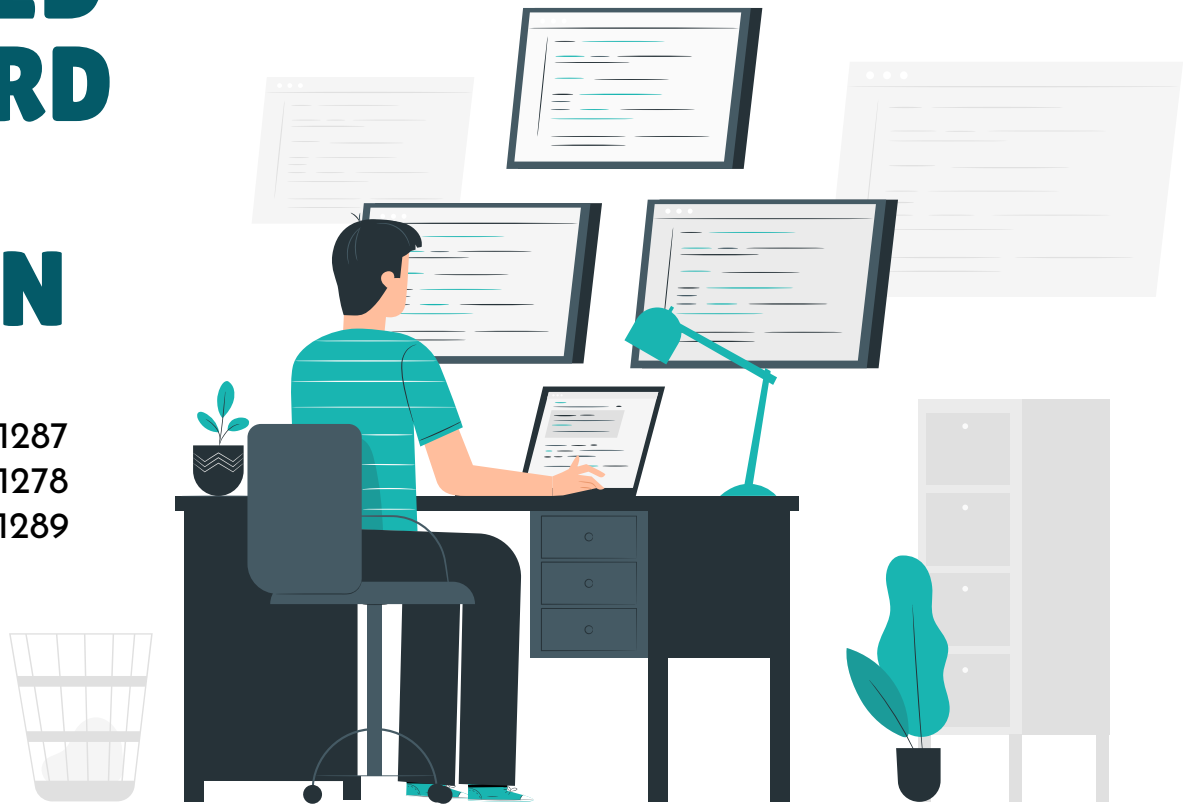
AUTOMATED CREDIT CARD FRAUD DETECTION

M.Vrishin Reddy	19B81A1287
P.Pothan Prathap	19B81A1278
M.Rohith Reddy	19B81A1289

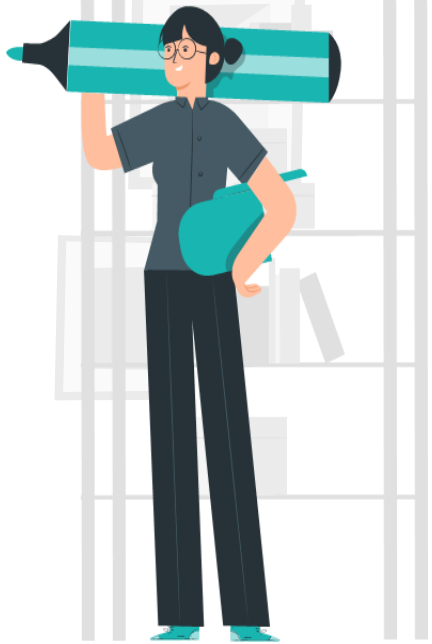


Internal Guide:
Dr.Bipin Bihari Jayasingh

20th october 2022

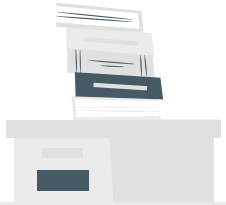
ABSTRACT

Everyone in the today’s generation is using the internet banking system for cash transfers, and the credit cards, debit cards are used for the shopping and online payments. Credit card fraud happens in a vey small percentage , yet the financial losses can be substantial . This needs the development of Automated Fraud Detection System(FDS). Credit card online transactions are the most commonly and widely used now-a-days. But with the exponential increase in the internet usage, the illegal transactions and illegal attacks are occurring in the financial industry. Without the knowledge of the authenticated user, unauthorized credit and debit card transactions, credit card theft and several other such fraudulent activities are alarming the world governments, clients and banking sector. The financial fraud detection systems can identify unusual attacks and unauthorized access. These fraud detection mechanisms are constantly updated by financial institutions. In order to prevent these illegal transactions and illegal attacks, we can use the technology such as Machine Learning, Deep learning techniques and try to predict the fraud transactions so that we could prevent the attacks and try to make the transactions done secured and efficient. In our project, we would like to use a financial fraud detection scheme using deep learning algorithm for the large volume of the data taken from the different sources from internet. It is observed that this methodology works very efficient for the fraud detection for the large data sets as RF is suitable for the small amount of data, LR results in overfitting.



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.
- Types of Frauds:- The credit-card fraud is divided into two types;
- The online credit card fraud (or no card present fraud) and (ii) The offline credit card fraud (card present fraud) The online credit-card fraud (also known as cyber credit card fraud) is committed with no presence of a credit-card but instead, the use of a credit-card information to make electronic purchase for goods and services on the internet.
- The offline credit-card fraud is committed with the presence of a credit-card which in most cases have been stolen or fake and thereby used at a local store or a physical location for the purchase or some goods or services.
- In order to prevent such fraudulent activities ,An automated detecting system is required to detect such activities and improvise such transactions to the users.



LITERATURE SURVEY

1. Introduction

Credit card fraud is a wide-ranging term for theft and fraud committed using or involving a payment card, such as a credit card or debit card, as a fraudulent source of funds in a transaction. The purpose of card fraud is to obtain goods or services, or cash, by unlawfully using another person's card or card information. Credit card fraud is a serious problem that costs card issuers, merchants, and consumers billions of dollars each year. In the United States alone, credit card fraud costs card issuers an estimated $8.6 billion annually. Merchants also incur significant costs, including chargebacks, fees, and the time needed to investigate and resolve fraud disputes. In addition, card fraud can cause consumers to lose confidence in the safety of electronic commerce and may deter them from shopping online.

2. Types of Credit Card Fraud

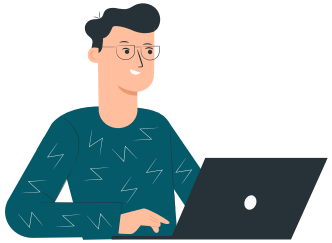
There are many different types of credit card fraud, but they can generally be divided into two broad categories:

- **Card-Present Fraud:** This type of fraud occurs when a thief uses a stolen or counterfeit credit card to make a purchase in person, over the phone, or online. Card-present fraud can also occur when a thief uses a stolen credit card number to make a purchase without the card physically present.
- **Card-Not-Present Fraud:** This type of fraud occurs when a thief uses a stolen credit card number to make a purchase without the card physically present. Card-not-present fraud can occur online, over the phone, or in person if the thief has access to the credit card number but not the physical card.

3. Credit Card Fraud Detection Methods

There are many different methods that can be used to detect credit card fraud. Some of the most common methods include:

- **Data Analysis**: Data analysis is a process of looking at data to find trends and patterns that can be used to detect fraud. Data analysis can be used to detect both card-present and card-not-present fraud.
- **Transaction Monitoring**: Transaction monitoring is a process of monitoring credit card transactions for suspicious activity. Transaction monitoring can be used to detect both card-present and card-not-present fraud.
- **Fraud scoring**: Fraud scoring is a process of assigning a score to each credit card transaction based on a set of fraud indicators. Transactions with a high fraud score are more likely to be fraudulent and can be flagged for further review.
- **Velocity checks**: Velocity checks are a process of monitoring the number of credit card transactions made in a given period of time.



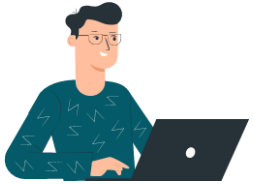
Problem Statement and Proposed Solution

Problem Statement

How can we prevent fraud transactions by using an automated system by different techniques.

Proposed Solution

The proposed solution for an automated credit card fraud detection system is to use machine learning algorithms to detect fraudulent activity. This can be done by training a machine learning model on historical data of credit card transactions. The model can then be used to predict whether a new transaction is likely to be fraudulent and here we discuss the different algorithms and its performance analysis which gives a better way of utilizing the algorithms to detect the fraud in a transaction.



MOTIVATION & SCOPE

- **Motivation**:

The use of machine learning in fraud detection has been an interesting topic now days. A credit card fraud detection algorithm consists in identifying those transactions with a high probability of being fraud, based on historical fraud patterns. Machine learning, having three types, from that also the supervised and hybrid approach is more suitable for fraud detection.

- **Scope**:

Predictability -

Data that can be used to detect fraud patterns to determine future probabilities and trends. Although predictive analytics won't reveal what type of fraud will happen, it will point to what might happen, with an acceptable degree of reliability.

The Credit Card Fraud Detection Problem includes modelling past credit card transactions with the data of the ones that turned out to be fraud. This model is then used to recognize whether a new transaction is fraudulent or not.

CONTRIBUTION

- Requirement Gathering : M.Rohith Reddy
- Data Modelling : P.Pothan Prathap
- Date Pre-Processing :M.Vrishin Reddy

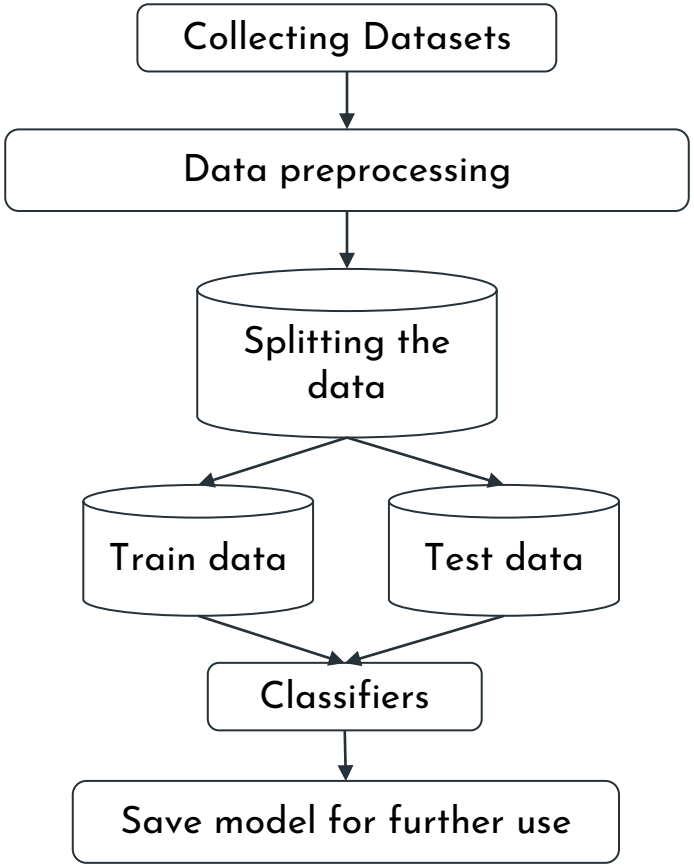
TECHNOLOGY STACK

Hardware Requirements

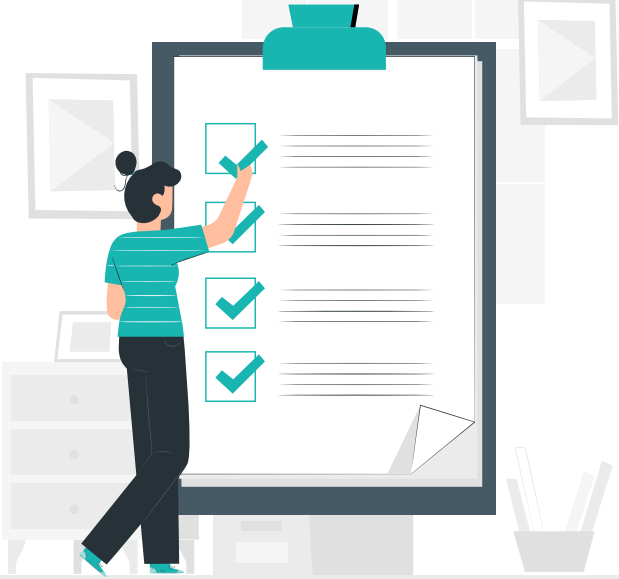
- i-5 or above
- CPU : quad core or above
- Gpu:2GB OR ABOVE
- RAM:8gb or above
- Operating System: Any OS of 64bit

Software Requirements

- **Programming language :** Python
  Python libraries:
  Sklearn (scikit-learn)
  Numpy
  Pandas
  Matplotlib
  Seaborn
- **Platform:**
  Google - Colab



Phases of Development



ALGORITHMS INVOLVED

LOGISTIC REGRESSION

- uses non-linear relationship.
- supervised learning algorithm, which means that it requires a labeled dataset in order to learn.
- The logistic regression algorithm is used to predict a binary outcome (1 or 0, true or false, yes or no).
- The logistic regression algorithm is coded in Python using the scikit-learn library.

DECISION TREE

- Support tool that uses a tree-like graph or model.
- Type of supervised learning algorithm that are used for both classification and regression tasks.
- The algorithm works by splitting the training data into multiple subsets, and then making predictions based on the features in each subset.
- The tree is constructed by starting at the root node, and then splitting the data into two subsets based on a certain criterion. The process is then repeated for each of the child nodes, until the tree is complete.

SUPPORT VECTOR MACHINE

- SVM is a supervised machine learning algorithm that can be used for both classification and regression tasks.
- The algorithm works by finding a hyperplane that best separates the data into classes.
- SVM is a powerful tool for both linear and non-linear classification.
- It is a non-probabilistic linear classifier.

NAIVE BAYES

- Naive Bayes algorithm is a simple supervised machine learning algorithm that is used for classification tasks.
- The algorithm is based on the Bayes theorem and assumes that the features in the data are independent of each other.
- This assumption simplifies the calculation of the probabilities, which is why the algorithm is called "naive".
- The Naive Bayes algorithm is trained on a data set and then makes predictions on new data. The predictions are based on the probabilities of the data set.

RANDOM FOREST

- It is a type of supervised learning algorithm that is used to create a model that predicts the value of a target variable by learning simple decision rules from data.
- Random forest is an ensemble learning algorithm that is used for classification and regression.
- It is a powerful tool for both classification and regression.
- It is a bagging technique that can be used to reduce the variance of a predictive model.
- It is also used to prevent overfitting.

FORMULAS FOR CLASSIFICATION METRICS

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

RESULT AND CONCLUSION

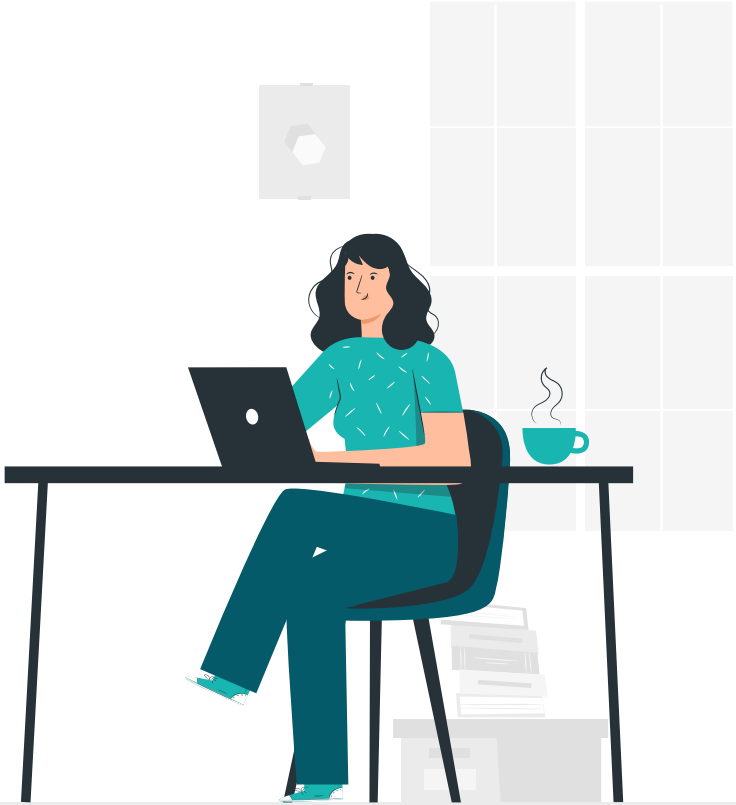
	Model	Accuracy	Recall	Precision	F1 Score
0	Decision Trees	0.999064	0.727941	0.697183	0.712230
1	Logistic Regression	0.998982	0.610294	0.709402	0.656126
2	Random Forest	0.999614	0.808824	0.940171	0.869565
3	Decision Trees	0.999064	0.727941	0.697183	0.712230
4	Support Vector Machine	0.999427	0.801471	0.832061	0.816479
5	Naive Bayes	0.993048	0.661765	0.141066	0.232558

We have tested our dataset consisting of 2L above transactions on various classifiers and the best accuracy was given by Random Forest model , which is 99.78% . We have chosen this classifier for fraud detection and deployed it in the google colab platform

Future Enhancement

01 Feature to detect the live based fraud detection

02 Enhancing the data sets with the information from real-time data to increase the real-time functionality of our model.



THANK YOU!