**PROJECT REPORT ON**

**Credit Card Fraud Detection**

Submitted by:

* P PREM SAI (20BCD7058)
* K ANURAG (20BCE7353)
* P SAI DEEPAK (20BCE7387)
* S PRATHEEK (20BCI7078)

UNDER THE GUIDANCE OF:

Dr. MONALI BORDOLOI

DEPT OF SCOPE

CERTIFICATE

This is to certify that the project is titled Credit card fraud detection. This project is submitted by P. Prem Sai, K. Anurag, P. Sai Deepak, S. Pratheek in fulfilment of the requirements for B.Tech. This project was an authentic work done by them under my supervision and guidance.

Date: 13th may, 2023

Monali Bordoloi

VIT-AP

ACKNOWLEDGEMENT

*We take this opportunity to express our profound gratitude and deep regards to my teacher Prof. Monali Bordoloi for her exemplary guidance, monitoring and constant encouragement throughout the course of this project. The blessing, help and guidance given by her time to time shall carry us a long way in the journey of life on which we are about to embark.*

ABSTRACT

A credit card is a thin handy plastic card that contains identification information such as a signature or picture, and authorizes the person named on it to charge purchases or services to his account - charges for which he will be billed periodically. Today, the information on the card is read by automated teller machines (ATMs), store readers, bank and is also used in online internet banking system. They have a unique card number which is of utmost importance. Its security relies on the physical security of the plastic card as well as the privacy of the credit card number. Credit cards are now the most preferred way for customers to transact either offline or online. There are a number of reasons, due to which consumers are slowly shifting from debit card transactions to credit cards, especially in developing countries like India.

Tie up credit cards with online and offline merchants, especially during festive seasons like Diwali, Eid, and Christmas, to offer further discounts on transactions. Several online merchants run their own promotional campaigns, which are tied up with credit cards—for example, Amazon Prime day, which happens at least once a year. Immediate needs can be fulfilled (for example, medical emergencies, lifetime events, etc.) quickly instead of having sufficient account balance for the same. Most credit cards offer 0% EMI options, so it makes it all the more worth pursuing this goal.

**Table of contents** Page No.

Abstract…….…………………………………………………………………… 4

Introduction..…………………………………………………………………. 6

Literature survey…..……………………………………………………….. 7

Proposed model………………………………….………………………… 8

Results………………………………………………...........………………… 9

Conclusion……....…………………………………………………………….. 12

References……………………………………………………………………… 13

Code………………………..……………………………………………………… 14

INTRODUCTION

A credit card is a thin handy plastic card that contains identification information such as a signature or picture, and authorizes the person named on it to charge purchases or services to his account - charges for which he will be billed periodically. Its security relies on the physical security of the plastic card as well as the privacy of the credit card number.

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 (Vector machines, random forest and decision trees) from that also the supervised and hybrid approach is more suitable for fraud detection.

To build credit card fraud detection system using machine learning algorithms. The major aim of this project is to perform a comprehensive review of different fraud detection methods like logistic regression and decision tree. Credit card fraud detection system is to identify suspicious events and report them to an analyst while letting normal transactions be automatically processed. We will make use of the Card Transactions dataset that contains a mix of fraud as well as non-fraudulent transactions.

LITERATURE SURVEY

Popat and Chaudhary supervised algorithms were presented Deep learning, Logistic Regression, Naive Bayesian, Support Vector Machine (SVM), Neural Network, K Nearest Neighbor, Data Mining, Decision Tree, Fuzzy logic-based System, and Genetic Algorithm are some of the techniques used.

Shiyang Xuan. For training the behavioral characteristics of credit card transactions, the Random Forest classifier was used. To assess the model's effectiveness, performance measures are computed. Using the Sliding-Window method, the transactions were aggregated into respective groups, i.e., some features from the window were extracted to find cardholder's behavioral patterns. Features such as the maximum amount, the minimum amount of a transaction, the average amount in the window, and even the time elapsed are available.

Sangeeta Mittal. To evaluate the underlying problems, some popular machine learning algorithms in the supervised and unsupervised categories were selected.

Siddhant. Bagga presented several techniques for determining whether a transaction is real or fraudulent Evaluated and compared.

Asha R B. Have proposed a deep learning-based method for detecting fraud in credit card transactions.

PROPOSED MODEL

**Random Forest**

Random forest is commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fuelled its adoption, as it handles both classification and regression problems.

**Logistic Regression**

Logistic regression, also called a logit model, is used to model dichotomous outcome variables. In the logit model the log odds of the outcome are modelled as a linear combination of the predictor variables.

**Decision Tree**

Decision Trees are a popular Data Mining technique that makes use of a tree- like structure to deliver consequences based on input decisions. One important property of decision trees is that it is used for both regression and classification. This type of classification method is capable of handling heterogeneous as well as missing data. Decision Trees are further capable of producing understandable rules. Furthermore, classifications can be performed without many computations.

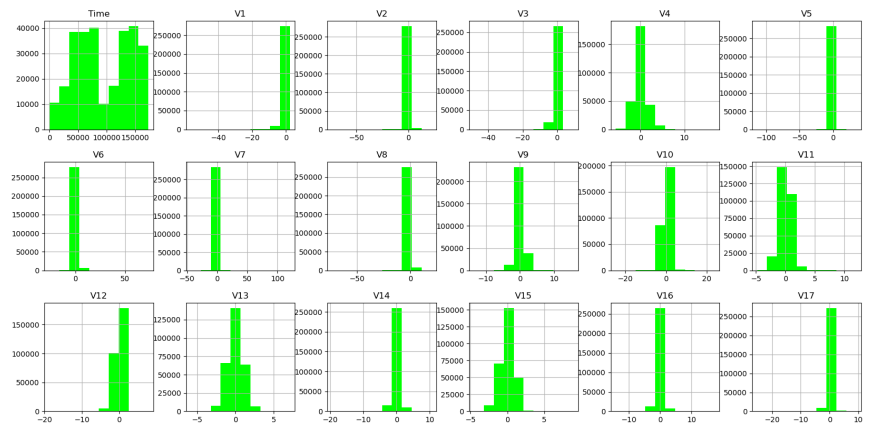
**Approach**

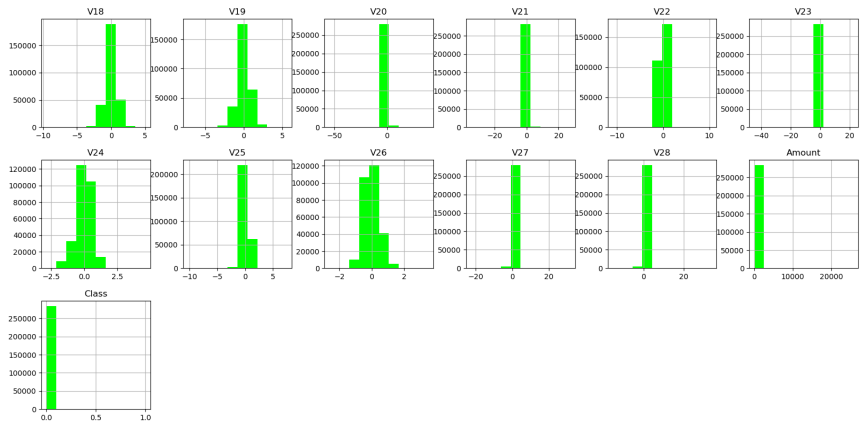
First, we load the data and check for the null values. Perform exploratory data analysis by plotting histograms, correlation heatmap and scatter plots between amount and time lapse. We build our models (Random Forest, Linear Regressor and Decision Tree. We test the data and calculate the accuracy of the model. We predict a new entry.

RESULTS

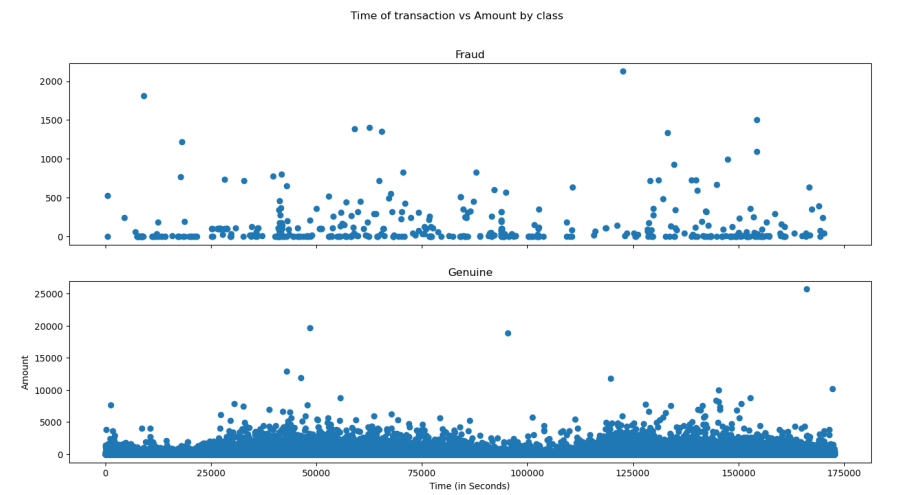
**Exploratory data analysis**

**Histograms**

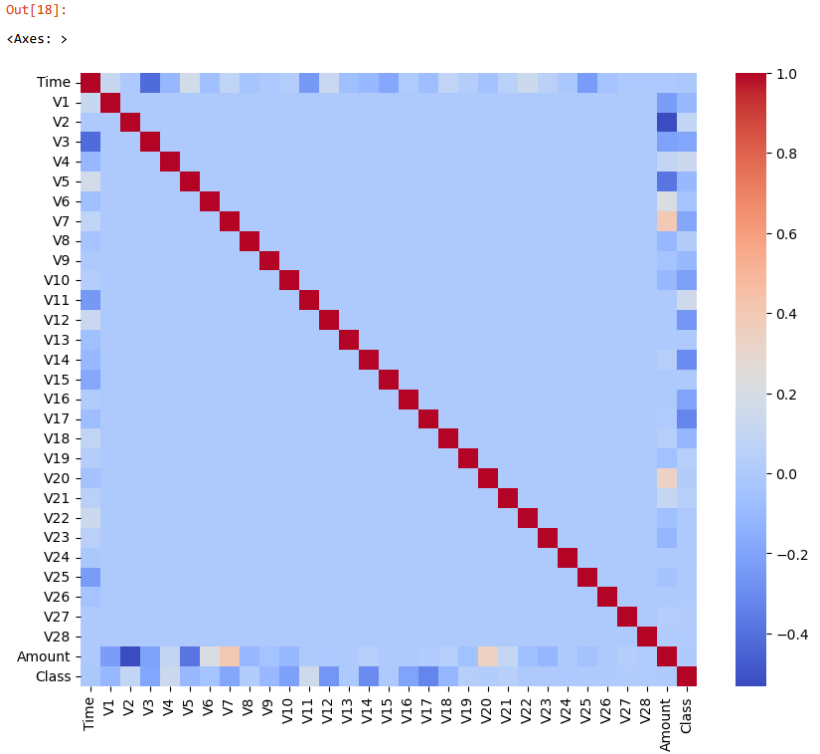




**Scatter plot**

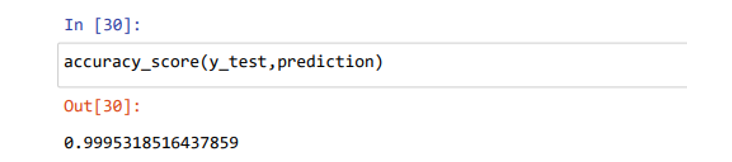


**Correlation matrix**

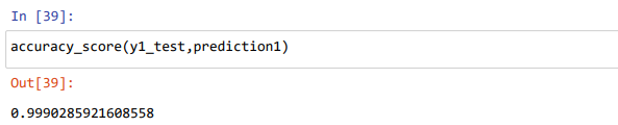


**Accuracy**

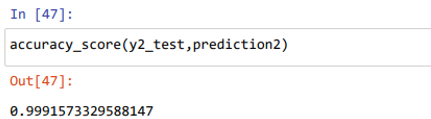
* **Accuracy score of Random Forest**

****

* **Accuracy score of Logistic Regressor**

****

* **Accuracy score of Decision Tree**

****

CONCLUSION

This method proves accurate in finding out the fraudulent transactions and minimizing the number of false alerts. The use of the above algorithms in credit card fraud detection system results in detecting or predicting the fraud probably in a very short span of time after the transactions has been made. This will eventually prevent the banks and customers from great losses and also will reduce risks. Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common detection methods and reviewed recent findings in this field. This project has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, explanation its implementation and experimentation results. While the algorithm does reach over 99.6% accuracy, this high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions

REFERENCES

* <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>
* <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00573-8>
* <https://www.ijitee.org/wp-content/uploads/papers/v10i6/C84000110321.pdf>
* <https://spd.group/machine-learning/credit-card-fraud-detection/>

CODE

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from pylab import rcParams

import warnings

warnings.filterwarnings('ignore')

data=pd.read\_csv(r'D:\UNIVERSITY\6TH SEM\DATA ANALYTICS\creditcard.csv')

data.head()

data.isnull().sum()

data.info()

data.describe().T.head()

print("Number of rows and columns",data.shape)

fruad\_cases=len(data[data['Class']==1])

print('Number of Fraud cases: ',fruad\_cases)

non\_fraud\_cases=len(data[data['Class']==0])

print('Number of Non Fruad cases: ',non\_fraud\_cases)

fraud=data[data['Class']==1]

genuine=data[data['Class']==0]

genuine.Amount.describe()

data.hist(figsize=(20,20),color='lime')

rcParams['figure.figsize'] = 16, 8

f,(ax1, ax2) = plt.subplots(2, 1, sharex=True)

f.suptitle('Time of transaction vs Amount by class')

ax1.scatter(fraud.Time, fraud.Amount)

ax1.set\_title('Fraud')

ax2.scatter(genuine.Time, genuine.Amount)

ax2.set\_title('Genuine')

plt.xlabel('Time (in Seconds)')

plt.ylabel('Amount')

plt.show()

plt.figure(figsize=(10,8))

corr=data.corr()

sns.heatmap(corr,cmap="coolwarm")

X=data.drop(['Class'],axis=1)

y=data['Class']

X\_train,X\_test,y\_train,y\_test=train\_test\_split(X,y,test\_size=0.30,random\_state=123)

from sklearn.ensemble import RandomForestClassifier

rfc=RandomForestClassifier()

model=rfc.fit(X\_train,y\_train)

prediction=model.predict(X\_test)

from sklearn.metrics import accuracy\_score

accuracy\_score(y\_test,prediction)

from sklearn.linear\_model import LogisticRegression

X1=data.drop(['Class'],axis=1)

y1=data['Class']

X1\_train,X1\_test,y1\_train,y1\_test=train\_test\_split(X1,y1,test\_size=0.3,random\_state=123)

lr=LogisticRegression()

model1=lr.fit(X1\_train,y1\_train)

prediction1=model1.predict(X1\_test)

accuracy\_score(y1\_test,prediction1)

from sklearn.tree import DecisionTreeRegressor

X2=data.drop(['Class'],axis=1)

y2=data['Class']

dt=DecisionTreeRegressor()

X2\_train,X2\_test,y2\_train,y2\_test=train\_test\_split(X2,y2,test\_size=0.3,random\_state=123)

model2=dt.fit(X2\_train,y2\_train)

prediction2=model2.predict(X2\_test)

accuracy\_score(y2\_test,prediction2)

model.predict([[172796,-1.359807236,-1.0727897,1.536592738,2.492155224,-0.33892677,1.462387928,0.239438754,1.092490901,1.36394697,0.090794172,-0.551599533,-0.617800856,-0.991389847,-0.311169354,1.468176972,-0.470400525,0.084571242,1.025754058,0.40329596,0.053212098,-1.018306778,0.277429576,-0.11593391,1.066928075,0.128539051,-0.189395444,0.133285677,-0.021543753,153.62]])