ONLINE PAYMENTS FRAUD DETECTION USING MACHINE LEARNING

AN INDUSTRIAL ORIENTED UG PHASE-1 REPORT

Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted By

UPPULA DIVYA 19UK1A05F5

VEMUNOORI RAMANA 19UK1A05F4

DEVA NAGESH 19UK1A05K0

VEMURU JAGADEESHWARI 19UK1A05G3

Under the guidance of

Mr. G.RAMESH

(Associate Professor)



DEPARTMENT OF COMPUTER SCIENCE ENGINEERING VAAGDEVI ENGINEEERING COLLEGE

Affiliated to JNTU, HYDERBAD BOLLIKUNTA, WARANGAL, (T.S)-506005 2019-2023

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING VAAGDEVI ENGINEEERING COLLEGE BOLLIKUNTA, WARANGAL, (T.S)-506005



CERTIFICATE

This is to certify that the UG Phase-1 entitled "ONLINE PAYMENTS FRAUD DETECTION USING MACHINE LEARNING" is being submitted by UPPULA DIVYA (19UK1A05F5), VEMUNOORI RAMANA (19UK1A05F4), DEVA NAGESH(19UK1A05K0), VEMURU JAGADEESHWARI(19UK1A05G3) in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2019-2023.

Project Guide Head of Department

Mr. G.RAMESH Dr. R. NAVEEN KUMAR

(Associate Professor) (Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. PRASAD RAO**, Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Phase-1 in the institute.

We extend our heartfelt thanks to **Dr. R. NAVEEN KUMAR**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Phase-1 and for their support in completing the UG Phase-1.

We express heartfelt thanks to the guide, **Mr. G. RAMESH**, Associate Professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Phase-1.

Finally, we express our sincere thanks and gratitude to our family members, friends for their encouragement and outpouring their knowledge and experiencing throughout this.

UPPULA DIVYA 19UK1A05F5

VEMUNOORI RAMANA 19UK1A05F4

DEVA NAGESH 19UK1A05K0

VEMURU JAGADEESHWARI 19UK1A05G3

ABSTRACT

In today's world, people depend on online payments for almost everything. Online transactions have their own merits like easy to use, feasibility, faster payments etc., but these kinds of transactions also have some demerits like fraud transactions, phishing, data loss, etc. With increase in online transactions, there is a constant threat for frauds and misleading transactions which can breach an individual's privacy. Hence, many commercial banks and insurance companies devoted millions of rupees to build a transaction detection system to prevent high risk transactions. We presented a machine learning - based transaction fraud detection model with some feature engineering. The algorithm can get experience; improve its stability and performance by processing as much as data possible. These algorithms can be used in the project that is online fraud transaction detection. In these, the dataset of certain transactions which is done online is taken. Then with the help of machine learning algorithms, we can find the unique data pattern or uncommon data patterns which will be useful to detect any fraud transactions. For the best results, the XGBoost algorithm will be used which is a cluster of decision trees. This algorithm is recently dominating this ML world. This algorithm has features like more accuracy and speed when compared to other ML algorithms.

Keywords – Fraud detection, Machine learning, Xgboost algorithm, classification, Data preprocessing, Prediction.

TABLE OF CONTENTS

| 1. INTRODUCTION | 6-7 |
|----------------------------------|---------|
| 1.1 Motivation | 6 |
| 1.2 Defination | 6 |
| 1.3 Objective of project | 7 |
| 1.4 Purpose | 7 |
| 2. PROBLEM STATEMENT | 8 |
| 3. LITERATURE SURVEY | 9-10 |
| 3.1 Exisisting problem | 9 |
| 3.2 Problem solution | 9-10 |
| 4. EXPERIMENTAL INVESTIGATION | NS11-14 |
| 4.1 Block diagram | 12 |
| 4.2 Architecture | 13 |
| 4.3 Software requirements | 13 |
| 4.4 Project flow | 14 |
| 5. DESIGN | 15 |
| 5.1 Dataset | 15 |
| 5.2 Use case diagram | 16 |
| 5.3 Flowchart | 16 |
| 6. CONCLUSION | 17 |
| 7. FUTURE SCOPE | 18 |
| LIST OF FIGURES | PAGE NO |
| Figure 1: Block Diagram | 12 |
| Figure 2: Architecture | 13 |
| Figure 3: Dataset table | 15 |
| Figure 4: Usecase diagram | 16 |
| Figure 5: Flowchart | 16 |

1.INTRODUCTION

1.1.MOTIVATION

In today's world, we are on the verge to become a cashless world. According to various surveys and researches, people performing online transactions has increased a lot, it's expected that in future years this will go on increasing. Now, while this might be exciting news, on the other-side fraudulent transactions are on the rise as well. Even due to various security systems being implemented, we still have a very high amount of money being lost due to fraudulent transactions. Online Fraud Transaction can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card-issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users to estimate, perceive or avoid objectionable behavior, which consists of fraud, intrusion, and defaulting. Most of the time, a person who has become a victim of such fraud doesn't have any idea about it until the very end.

1.2.DEFINITION

Online payment is the electronic transfer of funds via the internet, usually between a merchant and a consumer. These payments can be made in various ways, such as via credit and debit cards, banking apps or web pages. Exactly which method of online payment you choose to offer and accept will depend on the specifics of your business and the preferences of your target market.

Online payments are payments that are initiated over the internet for goods or services purchased either online or offline. Common methods to facilitate this include:

- Bank Debits via online mandate (often referred to a Direct debit which is the terminology we'll use in this guide)
- Bank transfers (also referred to as wire transfers)
- Online credit or debit card transactions.
- Digital wallet payments (such as PayPal)

OBJECTIVE OF PROJECT:

- You will able to Know fundamental concepts and techniques used for machine learning.
- You will Gain a broad understanding of data.
- You will Have knowledge of pre-processing the data/transformation techniques and some visualization concepts before building the model.
- You will Learn how to build a machine learning model and tune it for better performance.
- You will Know how to evaluate the model and deploy it using flask.

1.4.PURPOSE

The Main Purpose this Guided Project mainly focuses on applying a machine-learning algorithm for online payment fraud with machine learning, we need to train a machine learning model for classifying fraudulent and non-fraudulent payments. For this, we need a dataset containing information about online payment fraud, so that we can understand what type of transactions lead to fraud. For this task, we should collect dataset from Kaggle, which contains historical information about fraudulent transactions which can be used to detect fraud in online payments. online payment systems has helped a lot in the ease of payments. But, at the same time, it increased in payment frauds.

Minimising friction in your payment process saves you time and money and makes positive cash flow more likely. Therefore, it's important to choose payment collection methods that encourage prompt payment can be automated as possible.

PROBLEM STATEMENT

Necessary preventive measures can be taken to stop this abuse and the behavior of such fraudulent practices can be studied to minimize it and protect against similar occurrences in the future. In other words, this is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions for outnumber fraudulent ones. Also, the transaction patterns often change their statistical properties over time.

These are not the only challenges in the implementation of a real-world fraud detection system, however. In real world examples, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize. Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent. The investigators provide feedback to the automated system which is used to train and update the algorithm to eventually improve the fraud-detection performance over time. So, in this project, what we have tried is to create a Web App for the detection of such types of frauds with the help of Machine Learning.

LITERATURE SURVEY

3.1.EXISTING PROBLEM

With the growth of e-commerce websites & bank transactions people and financial companies rely on online services to carry out their transactions that have led to an exponential increase in the online payment frauds. Fraudulent payment transactions lead to a loss of huge amount of money. The design of an effective fraud detection system is necessary in order to reduce the losses incurred by the customers and financial companies. Research has been done on many models and methods to prevent and detect online payments frauds. Some payments fraud transaction datasets contain the problem of imbalance in datasets. A good fraud detection system should be able to identify the fraud transaction accurately and should make the detection possible in real-time transactions. Fraud detection can be divided into two groups: anomaly detection and misuse detection. Anomaly detection systems bring normal transaction to be trained and use techniques to determine novel frauds. Conversely, a misuse fraud detection system uses the labeled transaction as normal or fraud transaction to be trained in the database history. So, this misuse detection system entails a system of supervised learning and anomaly detection system a system of unsupervised learning. Fraudsters masquerade the normal behavior of customers and the fraud patterns are changing rapidly so the fraud detection system needs to constantly learn and update. Payments frauds can be broadly classified into three categories, that is, traditional card related frauds (application, stolen, account takeover, fake and counterfeit), merchant related frauds (merchant collusion and triangulation) and Internet frauds (site cloning, credit card generators and false merchant sites).

3.2.PROBLEM SOLUTION

We will be using classification algorithms such as Decision tree, Random forest, svm, and Extra tree classifier, xgboost Classifier .We will train and test the data with these algorithms. From this the best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.

we have used the Xgboost algorithm which also works based on the decision-making trees.

This algorithm has recently become popular dues to its advantages like fast, efficient, more accurate etc. the training proceeds iteratively, adding new trees that predict the residuals or errors of prior trees that are then combined with previous trees to make the final prediction. It's called gradient boosting because it uses a gradient descent algorithm to minimize the loss when adding new models. It basically classifies the transaction in only two states that are either frud or transation.

Xg boost Algorithm:

is a popular and efficient open-source implementation of the gradient boosted trees algorithm. Gradient boosting is a supervised learning algorithm, which attempts to accurately predict a target variable by combining the estimates of a set of simpler, weaker models.

When using gradient boosting for regression, the weak learners are regression trees, and each regression tree maps an input data point to one of its leafs that contains a continuous score. XGBoost minimizes a regularized (L1 and L2) objective function that combines a convex loss function (based on the difference between the predicted and target outputs) and a penalty term for model complexity (in other words, the regression tree functions). The training proceeds iteratively, adding new trees that predict the residuals or errors of prior trees that are then combined with previous trees to make the final prediction. It's called gradient boosting because it uses a gradient descent algorithm to minimize the loss when adding new models.

EXPERIMENTAL INVESTIGATIONS

Milestone 1: Data Collection

ML depends heavily on data, without data, a machine can't learn. It is the most crucial

aspect that makes algorithm training possible. In Machine Learning projects, we need a

training data set. It is the actual data set used to train the model for performing various

actions.

You can collect datasets from different open sources like kaggle.com, data.gov; UCI

machine learning repository etc. The dataset used for this project was obtained from

Kaggle.

Milestone 2: Data Pre-processing

Data Pre-processing includes the following main tasks

• Importing the libraries.

• Importing the dataset.

• Analyse the data.

• Taking care of Missing Data.

• Data Visualisation.

• Splitting Data into Train and Test

Milestone 3: Model Building

The model building process involves setting up ways of collecting data, understanding and

paying attention to what is important in the data to answer the questions you are asking,

finding a statistical, mathematical or a simulation model to gain understanding and make

predictions.

Model Building Includes:

- Import the model building libraries.
- Initialising the model.
- Training the model.
- Model Evaluation.
- Save the Model.

Milestone 4: Application Building

Create an HTML File.

- Build python code.
- Run the app in local browser.
- Show casting the prediction on UI.

4.1.BLOCK DIAGRAM

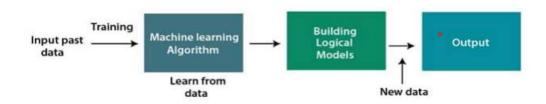
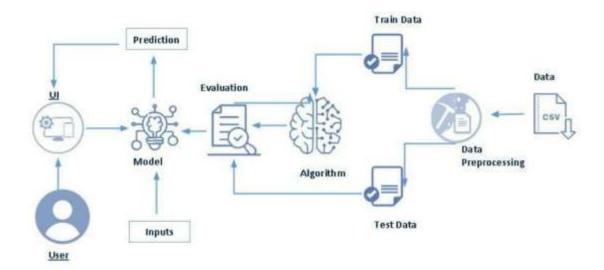


Figure 4: Block Diagram

4.2.ARCHITECTURE



4.3.SOFTWARE REQUIREMENTS

- > Python 3.9 or above:
 - Python is an interpreted high-level general-purpose programming language.
 - Python can be used on a server to create web applications.
- Visual Studio Code:
 - Visual studio code is a source-codeditor made by Microsoft for Windows, linux and macOS.
 - Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.

> Anaconda Environment

• The default environment base (path) is used because it consists of multiple libraries and modules.

Libraries

• Pandas and numpy, matplotlib, seaborn, and Algorithms etc..,

> Flask

- Flask is the module used for web framework.
- Flask provides you with tools, libraries and technologies that allow you to build a web application.

4.4.PROJECTFLOW:

- User interacts with the UI (User Interface) to upload the input features.
- Uploaded features/input is analysed by the model which is integrated.
- Once a model analyses the uploaded inputs, the prediction is showcased on the UI

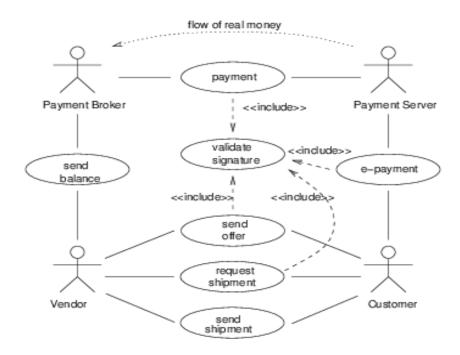
5.DESIGN

5.1.Dataset:

The dataset plays an important role in classifying the model. The dataset has been taken from the official Kaggle website, which is a well-informed data science organization. This dataset has details of millions of transactions out of which some of them are fraud transactions. This makes the development of the system more fluent and reliable. This dataset contains information on the rising risk of digital financial fraud, emphasizing the difficulty in obtaining such data. The main technical challenge it poses to predicting fraud is the highly imbalanced distribution between positive and negative classes in 6 million rows of data. The parameters of this dataset are Transaction type, amount, name Orig, oldbalance Org, newbalance Orig, name Dest, oldbalance Dest, newbalanceDest.

| variables | Description | Туре |
|----------------------|---|-------------|
| Transaction type | It states the type of the transaction | Categorical |
| Amount | Transaction amount | Numerical |
| Name-origin | Senders unique id | ID |
| Dest-Origin | Receivers unique id | ID |
| Old-balance- org | Senders balance before transaction | Numercial |
| New-balance -org | Senders balance after transaction | Numerical |
| Old-balance- dest | Receivers balance before transaction | Numercial |
| New-Balance -dest | Receivers balance after transaction | Numerical |

5.2.USE CASE DIAGRAM



5.3.FLOWCHART



6.CONCLUSION

In UG Project Phase-1, we have worked on problem statement, literature survey and also done the experimental analyses & design which are required for the project to move forward. In experimental analysis we have discussed about the machine learning concepts and models used in the project. We also discussed about the flowcharts, use case diagram which are used in the project. Based on the experimental analysis we have designed the model for the project. Entire designing part is involved in UG Project Phase-1.

7.FUTURE SCOPE

UG Project Phase-2 is the extension of UG Project Phase-1. UG Project Phase-2 involves all the coding and implementation of the design which we have retrieved from UG Project Phase-1. All the implementation is done and conclusions will be retrieved in the phase-II. We will also work on the applications, advantages, and disadvantages of the project in this phase. Future scope of the project will be also discussed in the UG Project Phase-2.

ONLINE PAYMENTS FRAUD DETECTION USING MACHINE LEARNING

AN INDUSTRIAL ORIENTED UG PHASE-2 REPORT

Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted By

UPPULA DIVYA 19UK1A05F5

VEMUNOORI RAMANA 19UK1A05F4

DEVA NAGESH 19UK1A05K0

VEMURU JAGADEESHWARI 19UK1A05G3

Under the guidance of

Mr.G.RAMESH

(Associate Professor)



DEPARTMENT OF COMPUTER SCIENCE ENGINEERING VAAGDEVI ENGINEEERING COLLEGE

Affiliated to JNTU, HYDERBAD BOLLIKUNTA, WARANGAL, (T.S)-506005 2019-2023

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING VAAGDEVI ENGINEEERING COLLEGE BOLLIKUNTA, WARANGAL, (T.S)-506005



CERTIFICATE

This is to certify that the UG Phase-2 entitled "ONLINE PAYMENTS FRAUD DETECTION USING MACHINE LEARNING" is being submitted by UPPULA DIVYA (19UK1A05F5), VEMUNOORI RAMANA (19UK1A05F4), DEVA NAGESH(19UK1A05K0), VEMURU JAGADEESHWARI(19UK1A05G3) in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2019-2023.

Project Guide Head of Department

Mr. G.RAMESH Dr. R. NAVEEN KUMAR

(Associate Professor) (Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. PRASAD RAO**, Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Phase-2 in the institute.

We extend our heartfelt thanks to **Dr. R. NAVEEN KUMAR**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Phase-2.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Phase-2 and for their support in completing the UG Phase-2.

We express heartfelt thanks to the guide, **Mr. G. RAMESH**, Associate Professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Phase-2.

Finally, we express our sincere thanks and gratitude to our family members, friends for their encouragement and outpouring their knowledge and experiencing throughout this.

UPPULA DIVYA 19UK1A05F5

VEMUNOORI RAMANA 19UK1A05F4

DEVA NAGESH 19UK1A05K0

VEMURU JAGADEESHWARI 19UK1A05G3

1. INTRODUCTION

In today's world, we are on the way to become a cashless world. According to various surveys and researches, people performing the online transactions is increased a lot, it's expected that in future years this will go on increasing. Now, while this might be exciting news, on the other-side fraudulent transactions are on the rise as well. Even due to various security systems being implemented, we still have a very high amount of money being lost due to fraudulent transactions. Online Fraud Transaction can be defined as a case where a person uses someone else's credit card for personal reasons or for knowing a persons personal info, while the owner and the card issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of users to estimate, perceive or avoid objectionable behavior, which consists of fraud, intrusion, and defaulting.

The online payment systems has helped a lot in the ease of payments. But, at the same time, it increased in payment frauds. Online payment frauds can happen with anyone using any payment system, especially while making payments using a credit card / debit card. That is why detecting online payment fraud is very important for credit card companies to ensure that the customers are not getting charged for the products and services they never paid.

Most of the E-commerce sites runs on online payments the fraudsters are ready to get the information / personal data once if the fraudster is known the card CVV number or payment UPI-ID then the fraudsters are entering and knowing the personal data of an individual, Even if they know the card number they can predict the CVV number. Because there are many ways now-a-days to predict and various algorithms to predict this may leads to the losing the personal data of a individual without is concern.

2. CODE SNIPPETS

2.1 MODEL CODE

```
Laport mampy as np
Laport matplotlib.pyplot as plt
import seaborn as ans
from sclay import stats
from sklearn.engerocessing import LabelEncoder
from sklearn.engedel selection import train_test_split
from sklearn.engedel selection import train_test_split
from sklearn.engerocessing import county_store
from sklearn.engerocessing import accuracy_store
from sklearn.engeroce import perisjonTreeclassifier
from sklearn.engerics import classification_report, confusion_matrix
from sklearn.engeroce import perisjonTreeclassifier
from sklearn.engeroce import perisjonTreeclassifier
from sklearn.engeroce import strainTreesClassifier
from sklearn.engeroce import strainTreesClassifier
from sklearn.engeroce import strainTreesClassifier
from sklearn.engeroce import accuracy_score
import xghoest as xgb

[5] data = pd.read_csv(r'/content/drive/MyOrlve/Major_proj_Outset/PS_20174392719_L401286419457_logs.csv')

[6] from google.colab import drive
drive_mount('/content/drive')

Drive_already_mounted_at_/content/drive; to attempt_to_forcibly_remount_call_drive_mount('/content/drive', force_remount=True).
```

Figure 1: .ipynb code importing libraries & mounting dataset from Drive.

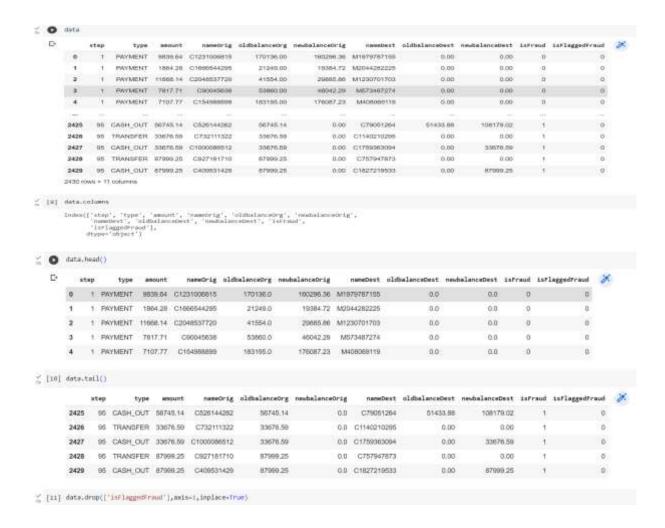


Figure 2: .ipynb code displaying few rows, columns & column names from the dataset.

```
data.info() #shows the descriptive statistics
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2430 entries, 0 to 2429
Data columns (total 10 columns):
     Column.
                     Non-Null Count Dtype
 0
     step
                     2430 non-null
                                     int64
     type
                     2430 non-null
                                     object
     amount
                     2430 non-null
                                      float64
     nameOrig
                     2430 non-null
     oldbalanceOrg
                     2430 non-null
                                      float64
     newbalanceOrig
                     2430 non-null
                                      float64
     nameDest
                     2430 non-null
                                     object
     oldbalanceDest 2430 non-null
                                      float64
     newbalanceDest
                     2430 non-null
                                     float64
     isFraud
                     2430 non-null
                                      int64
dtypes: float64(5), int64(2), object(3)
memory usage: 190.0+ KB
```

Figure 3: .ipynb code describe in detail info using info() method.



Figure 4: .ipynb code for heatmap shows 2 dimensional representation of dataset.

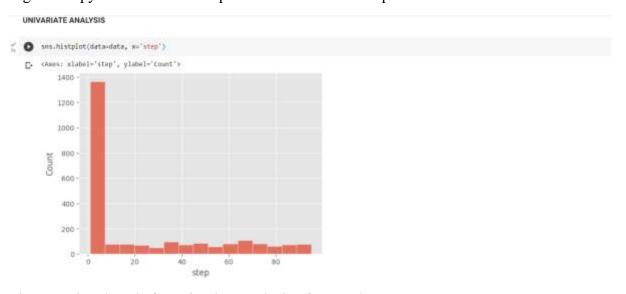


Figure 5: .ipynb code for univariate analysis of step column.

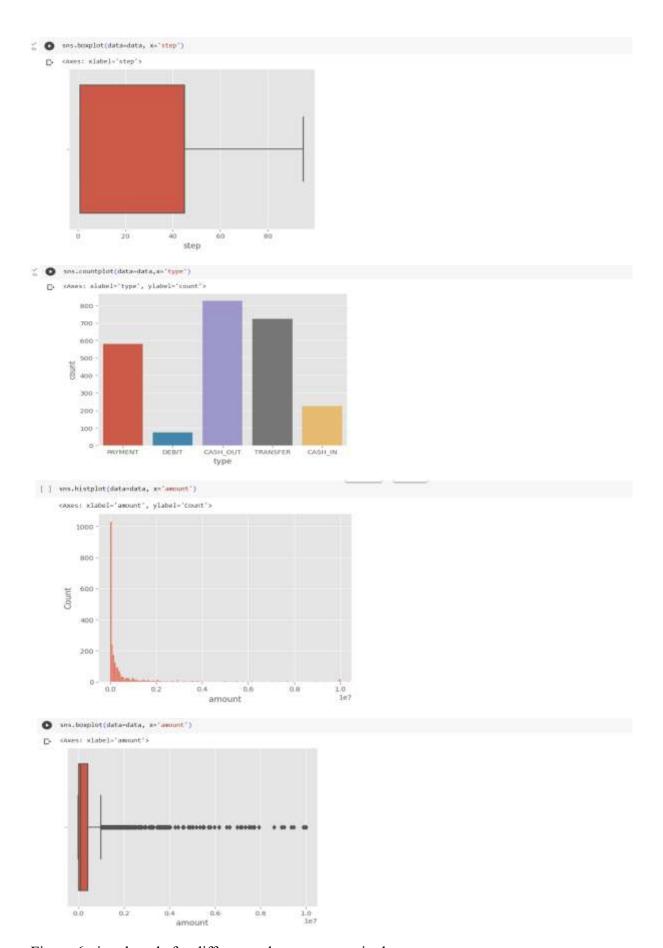


Figure 6: .ipynb code for different columns present in dataset.

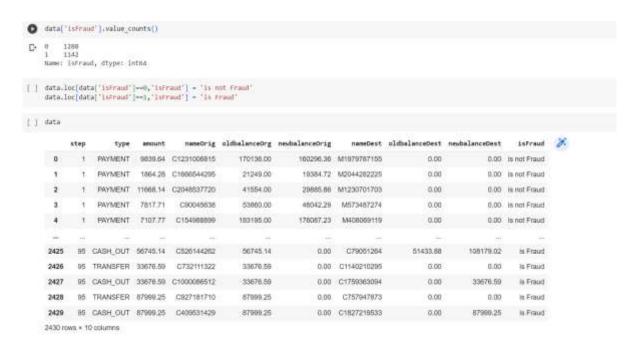
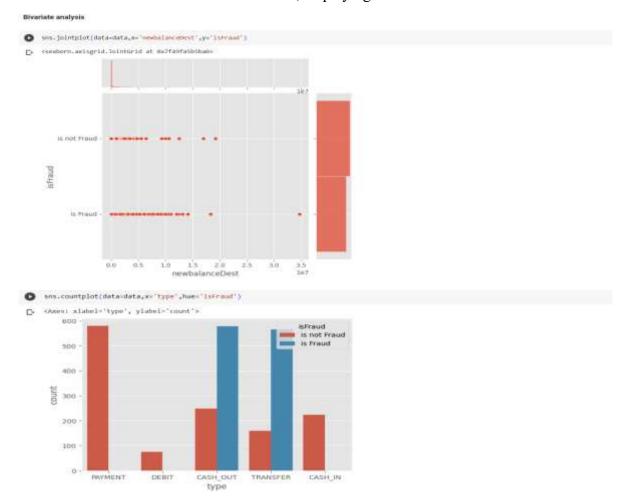


Figure 7: .ipynb code for count of fraud and non fraud transactions & Assigining is fraud=1 & is not fraud=0, displaying dataset.



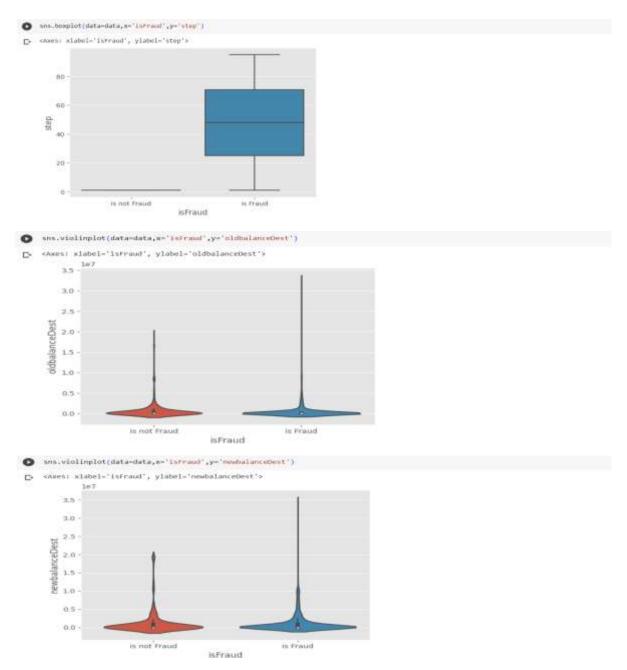


Figure 8: .ipynb code displaying Bi-variate analyssis gives relationship between each variable in dataset.

| | data.des | pribe(leclad | e-'all') | | | | | | | | | |
|---|----------|--------------|----------|---------------|-------------|---------------|---------------|--------------|----------------|----------------|--------------|----|
| > | | *149 | type | secunt | nameOrig | oldbalanceOrg | undalanceorig | nameDest | oldbelancedest | newbalanceDest | Ascount | 20 |
| | count | 2430.000000 | 2430 | 2.4900000+03 | 2410 | 2.4300006+03 | 2.4300006+03 | 2430 | 2,430000e+03 | 2,430000e+03 | 2410 | |
| | unique | 14301 | 0 | None | 2430 | trate | Nate | 1870 | 76474 | NaN | 2 | |
| | top | teate | GASH_OUT | None | C1231006818 | Plane | Teahs | IC1580650416 | hishi | MaN | is not froud | |
| | freq | 741074 | 827 | Nan | 1 | history | Nation | 26 | Testi | Nahi | 1088 | |
| | mean | 29.216046 | Nahi | 0.258361e+08 | PAIR | 9.848040e+05 | 4.392700e+00 | 74074 | 6.797246e+06 | 1.127025e+06 | None | |
| | shd | 29.933036 | Penery | 1.0038886+16 | teste | 2.082381e+08 | 1.H2097Ne+06 | Péople | 1.891192e=06 | 2.007401e=08 | Nore | |
| | maye | 1,000000 | Nunc | 8.7300000+00 | NoN | 0.000000e+00 | 0.0000000+00 | THIN | 0.000000w=00 | 0.00000004+00 | New | |
| | 25% | 1.000000 | Nun | 0.0184930+03 | Plate | 8.679630e+03 | 0.0000000+00 | Net | 0.00000004+00 | 0.000000e+00 | New | |
| | 50% | 1.000000 | NaN | 1.0505009+08 | NaN | 6.000250e+04 | 0.000000+00 | hore | 0.000000e=00 | 0.000000e+00 | Num | |
| | 75% | 45.000000 | Net | 4.00000000+00 | 741074 | 7,6002586+05 | 1.2478046+04 | Sum: | 3:099199e+05 | 8.666701e+05 | 1665 | |
| | mes | 96.000000 | None | 1.000000e+6T | Nah | 1.890000e+07 | 9.867287e+06 | New | 3.300000e+07 | 3.460000e+07 | heads | |

Figure 9: .ipynb code for descriptive analysis it describes the data.

Data Preprocessing

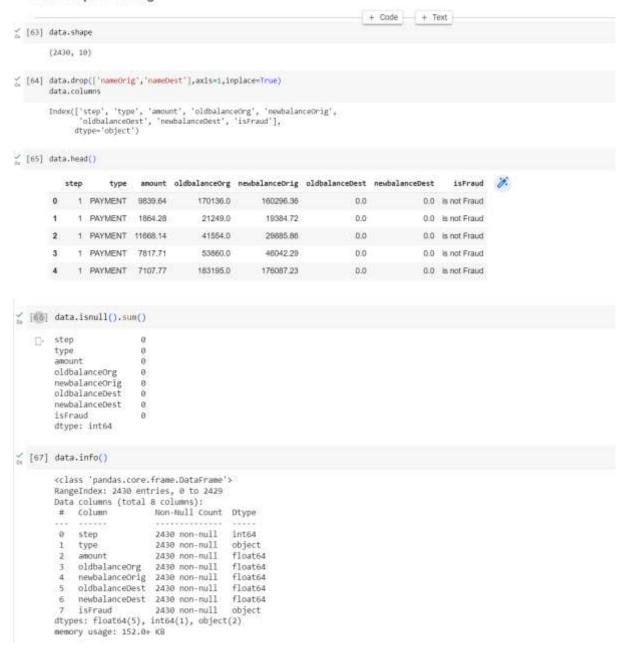


Figure 10: .ipynb code for Data preprocessing, Raw data to processing procedure.

· Remove the Outliers

```
from sitps (squer state princip, mean(data) (manuf())) princip, mean(data) (manuf())) princip, mean(data) (manuf())) princip, mean(data) (manuf())) princip, mean(data) (manuf(), manuf(), manuf
```

Figure 11: .ipynb code for removing outliers & transformation plot values.

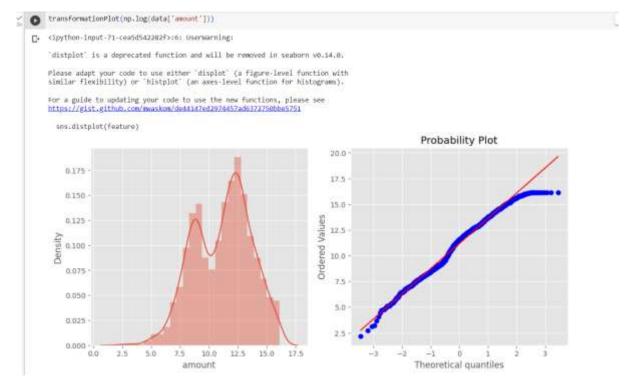


Figure 12: .ipynb code for transformation plot & graphs.

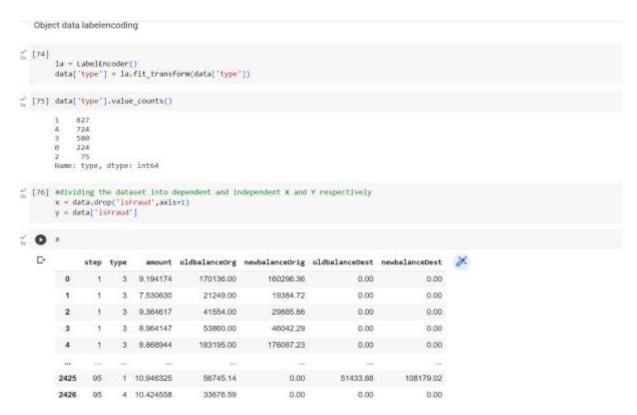


Figure 13: .ipynb code for object label encoding converts categorical values to numerical.

```
. 0
                is not Fraud
   C+
                is not Fraud
                is not Fraud
                is not Fraud
                is not Fraud
                    is Fraud
        2425
        2426
                    is Fraud
        2427
                    is Fraud
        2428
                    is Fraud
        2429
                    is Fraud
        Name: isFraud, Length: 2430, dtype: object
[79] #Splitting data into train and test
        x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=0,test_size=0.2)
[80] print(x_train.shape)
print(x_test.shape)
        print(y_test.shape)
       print(y_train.shape)
        (1944, 7)
        (486, 7)
        (486.)
        (1944,)
```

Figure 14: .ipynb code splitting data into train and test.

Model Building

Random Forest classifier



Figure 15: .ipynb code for Random Forest model.

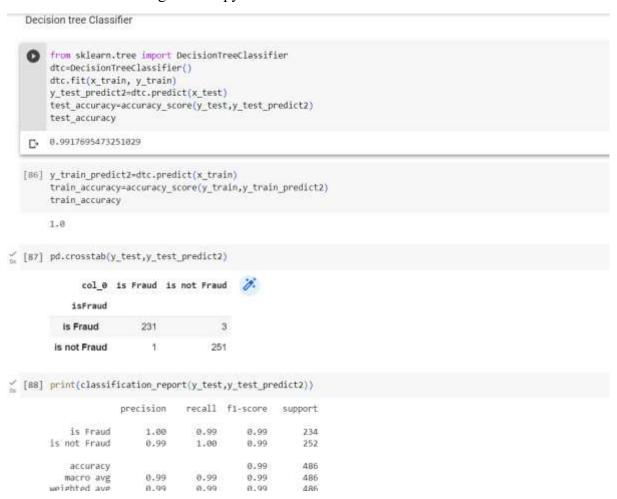


Figure 16: .ipynb code for Decesion tree classifier.



Figure 17: .ipynb code for extra trees classifier.

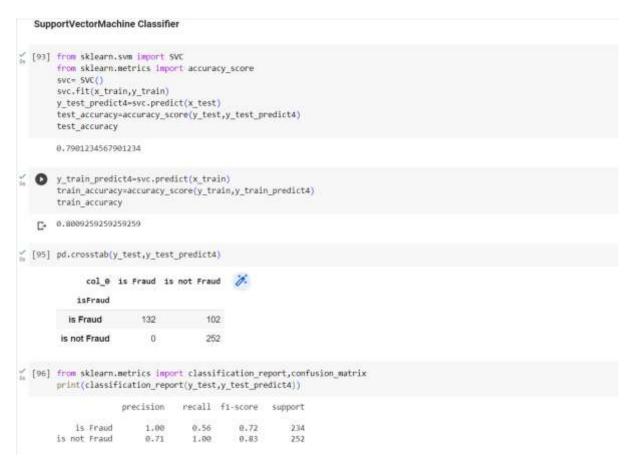


Figure 18: .ipynb code for support vector machine classifier.

Figure 19: .ipynb code for Label encoding converts categorical columns to numerical columns.

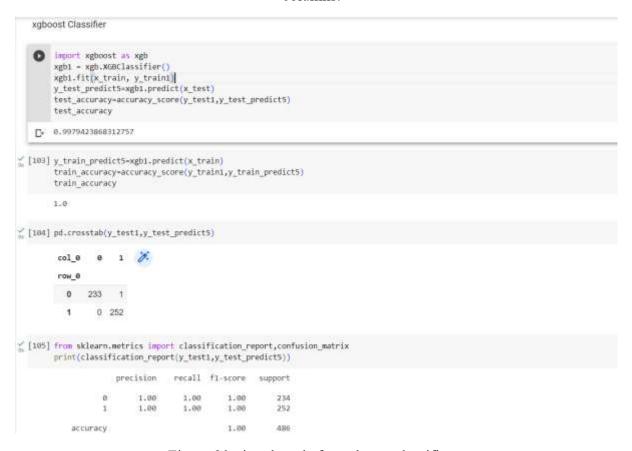


Figure 20: .ipynb code for xgboost classifier.

Figure 21: .ipynb code for comparing the models & accuracy of each model, importing pickle file(.py code).

Figure 22: .ipynb code for prediction & predicting by giving values.

2.2 HTML CODE AND PYTHON CODE

1. app.py code:

```
Spyder (Python 3.9)
 File Edit Search Source Run Debug Consoles Projects Tools View Help
 🗈 📴 👸 👂 🖂 🔘 🗓 🖟 🧸 🕒 🧸 🛊 🕒 🕒 🛊 🕒 🛊 🕒 🕒 🔞 🗗 🗡 🔞 Cliberi (anguel December) (color pay trail detector) (and colors of the color)
 C-(Josen)/Ragesh/Gravitateds/priline pay head detection/priline pay froud detection/field/app.py
    1 froe flask import Flask, render_template, request
2 import mampy as np
3 import pickle
40 import pandas as pd
           {\bf model} = {\tt pickle.load(open(r^*C; l/!sers|Wagesh)OneOrive|Desktap|anline|payments|f(ask|payments.pkl^*, 'rb'))}
           dup.route("/")
def about():
               return render_template('home.html')
          lings(route("/Nome")
def about1():
    return render_template('home.html')
           dup.route("/predict")
def homel():
                return render_template('predict.htm(')
           dupp.route("/pred", methods=['POST','GET'])
def predict():
             x = [[x for x in request.form.values()]]
print(x)
              x = np.array(x)
print(x.shape)
             print(x)
pred = model.predict(x)
print(pred(0))
return render_template('submit.html', prediction_text-str(pred))
           if __name__ -- __moin__":
spp.rum(debug=False)
```

Figure 23: .python code used for rendering all the HTML pages.

2. home.html

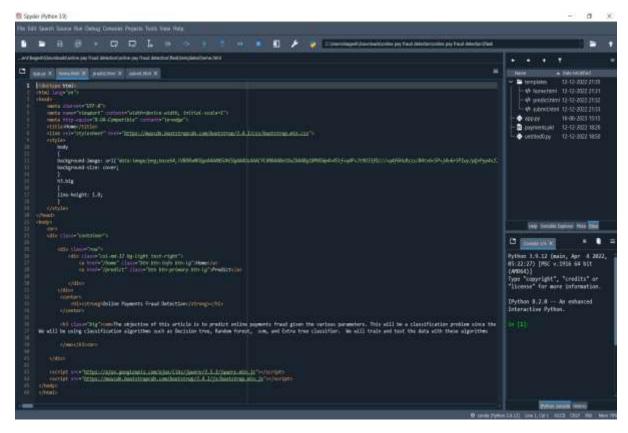


Figure 23: home.html page is the code for homepage of our web application.

3. predict.html:

```
Gets :hurset="VIII-8")
(title:Predict(/title)
(link rel="style:Deet" href="https://moscdn.ke
     bockground-laage: url("data:leage/prg;boseb4;lVBORDBROgeAAADBSDRCHgAADDAAAACYCAMAABDDDZAAANglBMTiglevHScf+q4P+JtSC33fD///+q4FeHc8szs/84Tv6v5P
bockground-size: cover;
     l
h3.big
```

Figure 24: predict.html page which predicts the output. By taking the inputs from user.

4. Submit.html

```
Spyder (Python 3.9)
 Re Edit Search Source Ruy Debug Consoles Projects Yooks View Help
     🗎 B B P 🖫 🗆 II, III 🥱 🛊 🐧 🕨 🙎 🔞 🖋 Characterpolithonius deleteration pay head deleteration pay head deleteration pay head deleteration pay head deleteration pay.
     publitenskietskoche pay fraud delactionkreine pay fraud delaction/fleaktionglebuikadenst.html
Stray X haralded X professor X married X
    1 | DOCTYPE html>
       chtml lang="en">
      <head>
<meta charset="UTF-8">
<title>Output</title>
       clink rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
       cstyles
               background-image: url("data:image/png;base64,ivBORw8KGgoAAAANSUhEUgAAAUsAAACYCAMAABatDuZAAABg1BHVENp4v45tf+q4P+Jt9
background-size: cover;
                h3.big
                line-height: 1.8;
      </style>
       (body)
                <div class="container">
                         </div>
      <hi><strong>Online Payments Fraud Detection</strong></hi><br/>cbr>
       The predicted fraud for the online payment is {{prediction_text}}
      </h3>
</div>
</body>
</html>
```

Figure 25: submit.html is a button when we enter values & click on submit button it displays a message associated with code.

3. CONCLUSION



Figure 26: Home page (which gives introduction to Online payments Fraud Detection)



Figure 27: Input page (which takes input from user)



Figure 28: Output page (Displays that the payment is fraud)



Figure 29: Input page (which takes input from user)



Figure 30: Output page (Displays that the payment is not fraud)

4. APPLICATIONS

The areas where this solution can be applied:

- Bank Transfers & Banking Applications.
- QR codes/UPI payments.
- Digital wallets like phone pe, paytm etc..,
- Swipping machines (card cvv).

5. ADVANTAGES

- 1. Improved Security: Online payment fraud detection projects employ advanced algorithms and techniques to identify and prevent fraudulent activities. This helps in enhancing the overall security of online transactions and protects both businesses and customers.
- **2. Real-Time Detection:** Online payment fraud detection systems can analyze transactions in real time, enabling the identification of suspicious patterns or behaviors instantly. This allows for immediate action to be taken, such as blocking a transaction or flagging it for manual review.
- **3. Cost Savings:** By implementing an effective fraud detection system, businesses can minimize financial losses due to fraudulent activities. Identifying and preventing fraudulent transactions early on can save significant amounts of money that would otherwise be lost.
- **4. Enhanced Customer Trust:** A robust fraud detection system reassures customers that their financial information is secure when making online payments. This helps to build trust and confidence in the business, leading to increased customer satisfaction and loyalty.
- **5. Scalability:** Online payment fraud detection systems can handle large volumes of transactions, making them scalable for businesses of different sizes. As the volume of online transactions increases, the system can adapt and accommodate the growing demands.

6. DISADVANTAGES

- 1. **False Positives:** One of the challenges in online payment fraud detection is the occurrence of false positives, where legitimate transactions are incorrectly flagged as fraudulent. This can inconvenience customers and lead to a loss of business if genuine transactions are blocked or delayed.
- 2. **Evolving Fraud Techniques:** Fraudsters are continually adapting their techniques to bypass detection systems. Keeping up with new and emerging fraud patterns and updating the fraud detection algorithms accordingly can be challenging.
- 3. **Privacy Concerns:** Online payment fraud detection projects involve the analysis of large amounts of personal and financial data. Ensuring the privacy and security of this sensitive information is crucial to prevent unauthorized access or data breaches.

6. FUTURE SCOPE

On our Dataset, we have applied Random Forest, Decision Tree, Xgboost Classifier, SVM, and Extra tree classifier, Xgboost has got the highest accuracy.

Enhancements that can be made in the future:

Online payment Fraud Transaction Detection System is basically an extension of the existing system. Using This system, the algorithms which we used to train the dataset and provide the appropriate output. In the long run, this system will be quite beneficial as it provides an efficient system to create a secure transaction system to analyse and detect fraudulent transactions. The Xgboost algorithm is a popular and efficient open-source implementation of the gradient boosted trees algorithm. Gradient boosting is a supervised learning algorithm, which attempts to accurately predict a target variable by combining the estimates of a set of simpler, weaker models. This accuracy can be increased further by providing a huge dataset for model training. The scope of this application is very far reaching. This system can be used to detect the features of fraud transactions in a dataset which is very well applicable in various sectors like banking, insurance, e-commerce, money transfer, bill payments, etc. This will indeed help to increase security.

7. BIBILOGRAPHY

- 1. K.Chaudhary, J.Yadav, "A review of fraud: A comparative study." decis. Support syst, vol 50, no3, pp.602-613,2011.
- 2. Katherine J. Barker, Jackie D'Amato, Paul Sheridon, 2008 "Credit card fraud: awareness and prevention", Journal+- of financial Crime, Vol. 15issue: 4,pp. 398-410.
- "CreditCard Fraud Detection Based on Transaction Be haviour -by John Richard D. Kho, Larry
 A. Vea" published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia,
 November 5- 8, 2017.
- 4. Customer Transaction Fraud Detection Using Xgboost Model -by Yixuan Zhang, Ziyi Wang, Jialiang Tong, Fengqiang Gao June, 2020.
- 5. Wang, M., Yu, J., & Ji, Z. (2018). Credit Fraud Risk Detection Based on XGBoost-LR Hybrid Model.
- Mishra, C. Ghorpade, "Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques" 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) pp. 1-5. IEEE.
- 7. https://thecleverprogrammer.com/2022/02/22/online-payments-fraud-detection-with-machine-learning/
- 8. https://www.geeksforgeeks.org/online-payment-fraud-detection-using-machine-learning-in-python/

9.HELP LINE

PROJECT EXCEUTION:

- STEP-1: Go to Google, search google colaboratory & launch.
- STEP-2: After launching of collab.
- STEP-3: Open "Major project .ipynb file."
- STEP-4: Then run all the cells.
- STEP-5: All the data preprocessing, training and testing, model building, accuracy of the model can be showcased.
- STEP-6: And a pickle file will be generated.
- STEP-7: Create a Folder named FLASK on the DESKTOP. Extract the pickle file into this Flask Folder.
- STEP-8: Extract all the html files (home.html, predict.html, submit.html) and python file(app.py) into the FLASK Folder.
- STEP-9: Then go back to ANACONDA NAVIGATOR and the launch the SPYDER.
- STEP-10: After launching Spyder, give the path of FLASK FOLDER which you have created on the DESKTOP.
- STEP-11: Open the app.py and html files present in the Flask Folder.
- STEP-12: After running of the app.py, open ANACONDA PROMPT and follow the below steps: cd File Path<> click enter python app.py< >click enter (We could see running of files).
- STEP-13: Then open BROWSER, at the URL area type >> localhost:5000.
- STEP-14: Home page of the project will be displayed.
- STEP-15: Click on Predict. Give the inputs then it will be predict fraud payment or not.