**Credit Card Fraud Detection**

The challenge is to recognize fraudulent credit card transactions so that the customers of credit card companies are not charged for items that they did not purchase.

**Main challenges involved in credit card fraud detection are:**

1.Enormous Data is processed every day and the model build must be fast enough to respond to the scam in time.

2.Imbalanced Data i.e most of the transactions (99.8%) are not fraudulent which makes it really hard for detecting the fraudulent ones

3.Data availability as the data is mostly private.

4.Misclassified Data can be another major issue, as not every fraudulent transaction is caught and reported.

5.Adaptive techniques used against the model by the scammers.

**How to tackle these challenges?**

1.The model used must be simple and fast enough to detect the anomaly and classify it as a fraudulent transaction as quickly as possible.

2.Imbalance can be dealt with by properly using some methods which we will talk about in the next paragraph

3 For protecting the privacy of the user the dimensionality of the data can be reduced.

4.A more trustworthy source must be taken which double-check the data, at least for training the model.

5.We can make the model simple and interpretable so that when the scammer adapts to it with just some tweaks we can have a new model up and running to deploy.

Before going to the code it is requested to work on a jupyter notebook. If not installed on your machine you can use Google colab.

You can download the dataset from this link

If the link is not working please go to this link and login to kaggle to download the dataset.

**Code** : Importing all the necessary Libraries

# import the necessary packages

Import numpy as np

Import pandas as pd

Import matplotlib.pyplot as plt

Import seaborn as sns

From matplotlib import gridspec

**Code** : Loading the Data

# Load the dataset from the csv file using pandas

# best way is to mount the drive on colab and

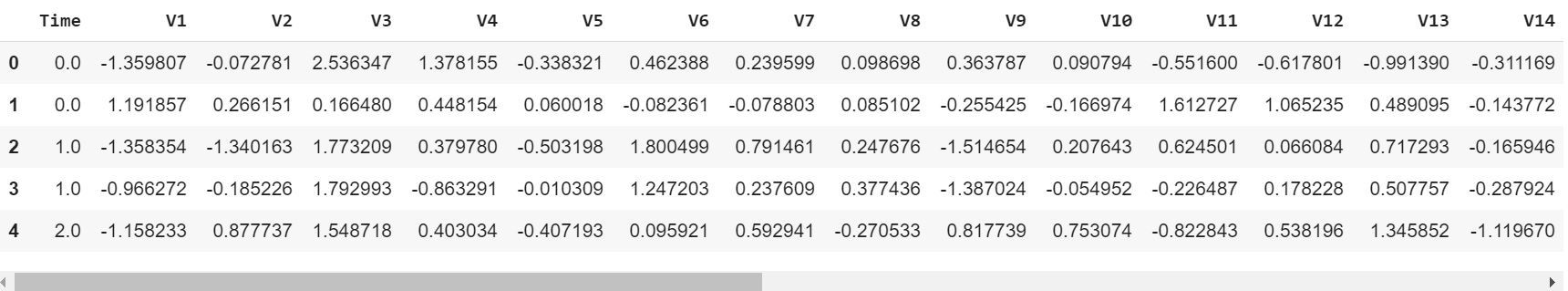
# copy the path for the csv file

Data =pd.read\_csv(“credit.csv”)

**Code** : Understanding the Data

#Grab a peek at the data

Data.head()



**Code** : Describing the Data

# Print the shape of the data

# data = data.sample(frac = 0.1, random\_state = 48)

print(data.shape)

print(data.describe())

**Output** :

(284807, 31)

Time V1 ... Amount Class

count 284807.000000 2.848070e+05 ... 284807.000000 284807.000000

mean 94813.859575 3.919560e-15 ... 88.349619 0.001727

std 47488.145955 1.958696e+00 ... 250.120109 0.041527

min 0.000000 -5.640751e+01 ... 0.000000 0.000000

25% 54201.500000 -9.203734e-01 ... 5.600000 0.000000

50% 84692.000000 1.810880e-02 ... 22.000000 0.000000

75% 139320.500000 1.315642e+00 ... 77.165000 0.000000

max 172792.000000 2.454930e+00 ... 25691.160000 1.000000