Credit Card Fraud Detection

The recent advances of e-commerce and e-payment systems have sparked an increase in financial fraud cases such as credit card fraud.It is therefore crucial to implement mechanisms that can detect the credit card fraud.Features of credit card fraud plays an important role when machine learning is used for credit card fraud detection,and they must be chosen properly.

MileStone 1: Problem Statement

The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be fraud. This model is then used to identify whether a new transaction is fraudulent or not. Our aim here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

MileStone 2: Data Processing

In this step we will find missing values or null values and even duplicate values if found any of these values we will remove them so that they are useful to find the data that is more efficient in performing the PCA(Principle Component Analysis).

MileStone 3: Data Visualization(Exploratory Data Analysis)

In this step we will understand the data in detail and we can understand each variable individually by calculating frequency counts, visualizing the distributions and plotting histograms,scatter graph etc.

In the process of exploring Data Analysis we perform,

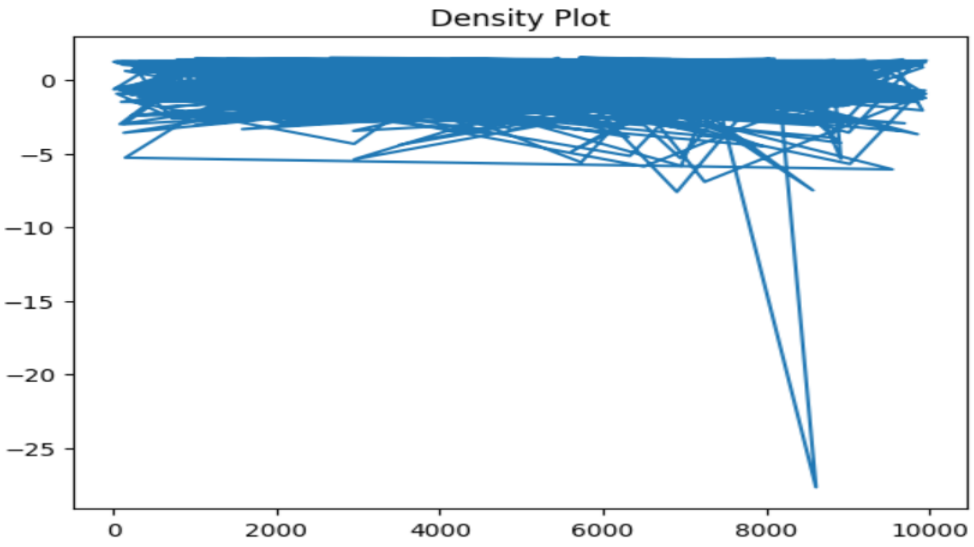
Univariate Analysis

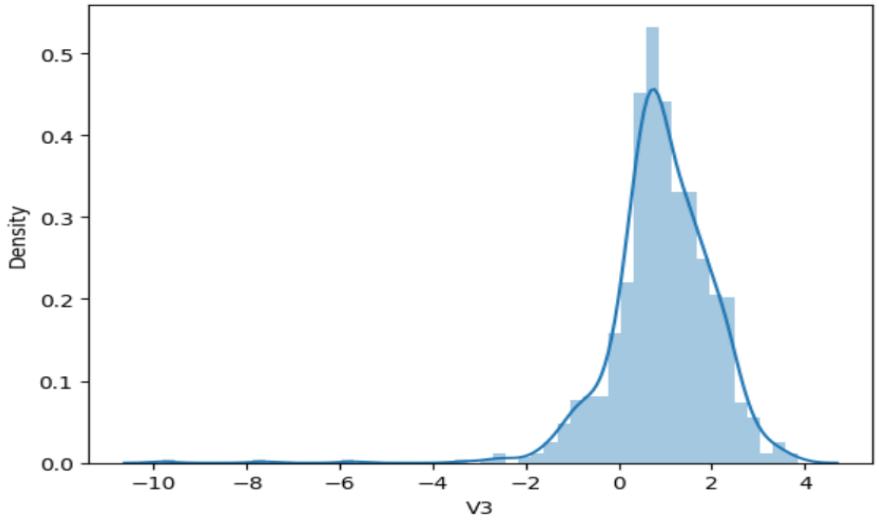
Bivariate Analysis

Multivariate Analysis

Univariate Analysis:

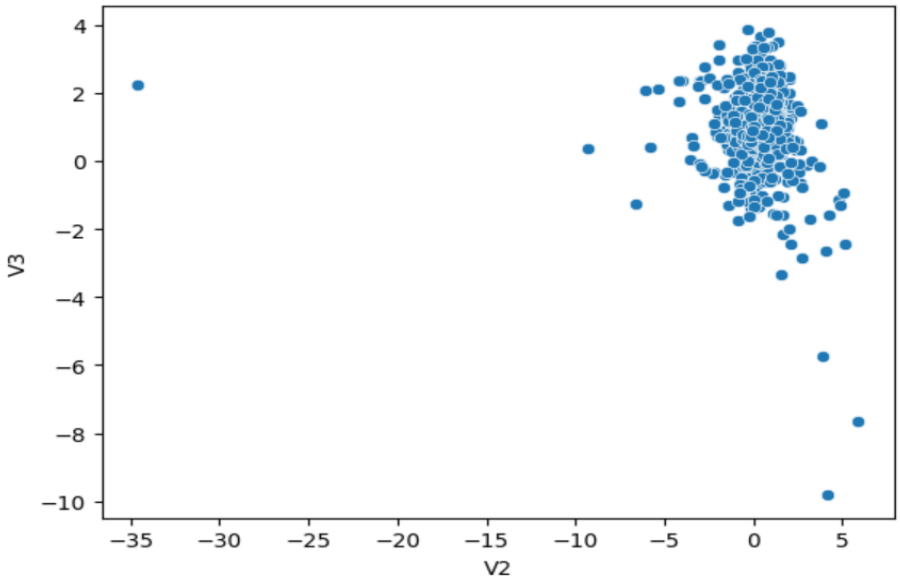
Here we have plotted a Density plot graph and aHistogram

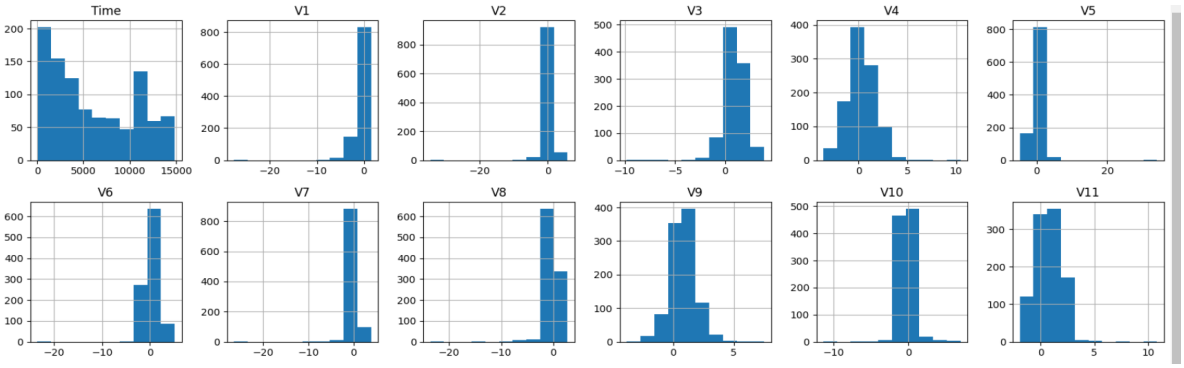


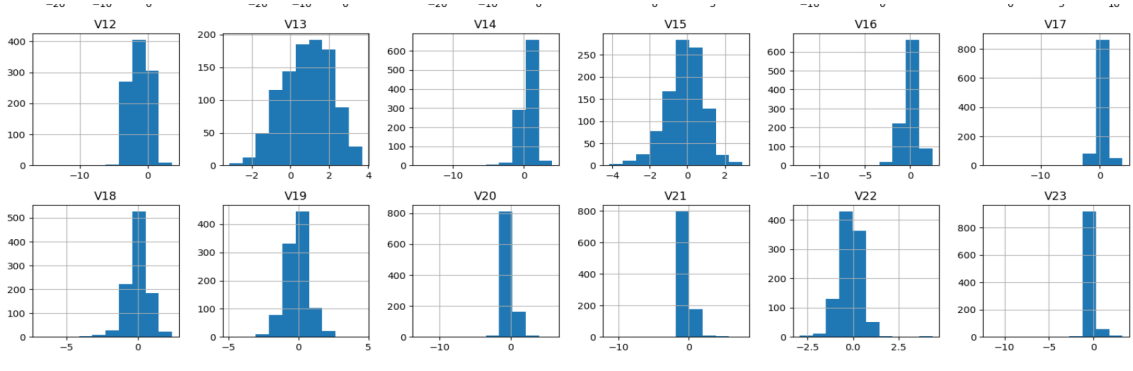


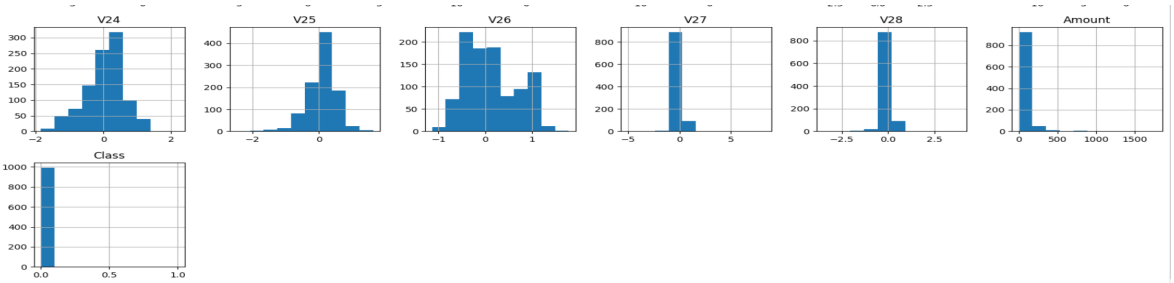
Bivariate Analysis:

In Bivariate Analysis analysis we will plot a scatter plot graph and Histograms for each value of V1,V2.....V28.


V3
V2


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

Multivariate Analysis:
In this we will plot a graph using multiple values for V1,.....V28.

MileStone 4: Model Building

In this step we perform Analysis by Splitting the data,Scaling the data and Testing the data. Here we used Random Forest,Decision Tree and Logistics Regression Techniques.Using these techniques we will find the r2_score ,mean squared error, and mean bias error.

MileStone 5: Deployment

In this stage we have used some algorithms such as Decision Tree,Random Forest and XGBoost Classifier to find out whether the fraud has taken place or not.We even performed Split,Test and train.
Hence ,We got a accuracy 99.96% with XGBoost mode.l