Online Fraud Transaction Detection using Machine Learning

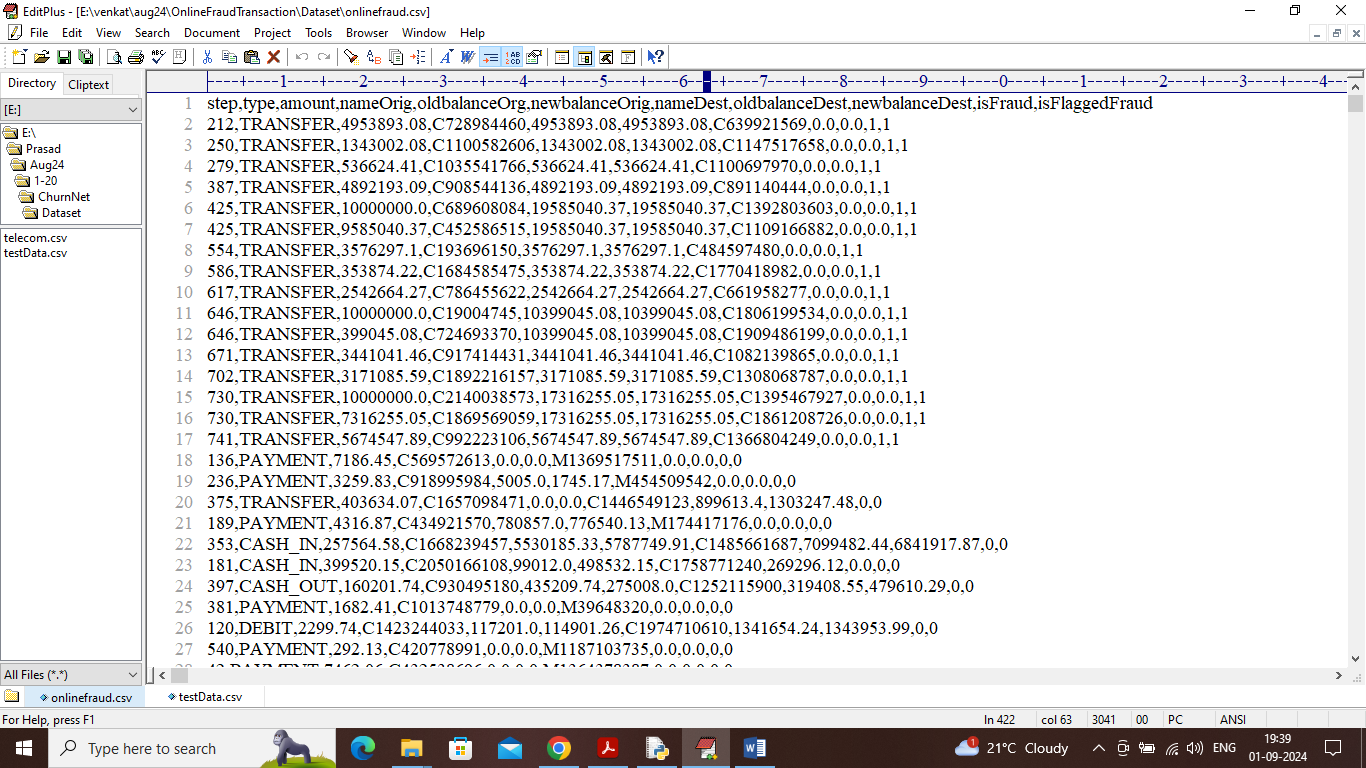
Growing technologies converting all manual works to virtual works such as online shopping or online transaction where users can shop online and can make payments online. This advantage leads to a problem called Fraud transactions where malicious attackers can perform phishing activities to generate fraud transaction. All banks investing billions of money to detect and prevent such frauds attacks but no algorithms are giving accurate prediction.

In propose work we are employing advance machine learning algorithm such as XGBOOST which has a group of estimators and decision trees for accurate prediction. This algorithm trained on fraud transaction dataset and manages to detect fraud transaction on test data with an accuracy of over 99%.

To enhance algorithm performance we have employed Principal Component Analysis (PCA) algorithm to select relevant features from the dataset and then selected features will be input to XGBOOST to train a model.

For training we are using 70% dataset and for testing we have utilize 30% dataset. We have experimented with existing Naïve Bayes algorithm and propose XGBOOST algorithm. Each algorithm performance is measured by using different metrics such as Confusion Matrix, Accuracy, Precision, Recall and FSCORE.

To train algorithms we have used below Fraud Transaction Dataset from KAGGLE repository and below screen showing dataset details



In above dataset screen first row represents dataset column names and remaining rows represents dataset values. By using above dataset will train and test all algorithms performance.

To implement this project we have implemented following modules

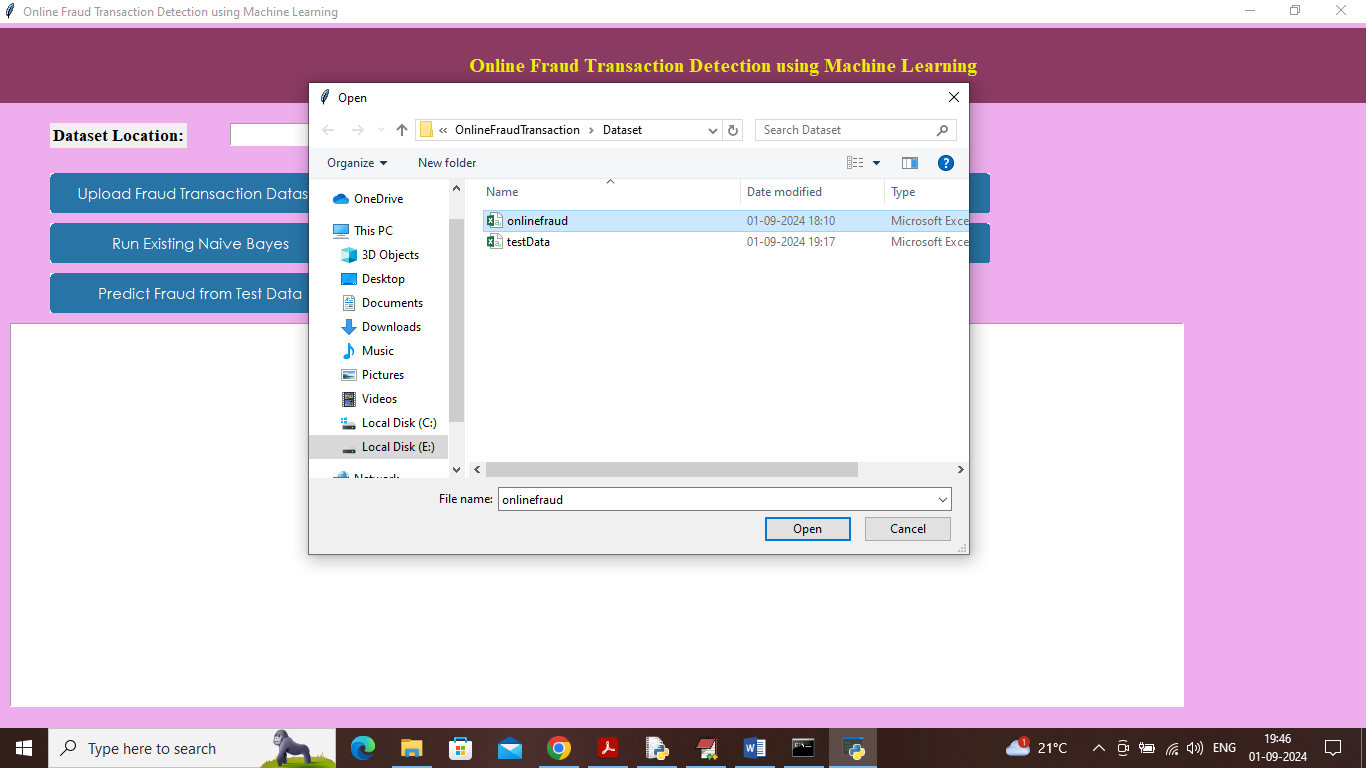
1. Upload Fraud Transaction Dataset: using this module we will upload and load dataset to application and then plot graph of real and fraud transaction
2. Pre-process Dataset: using this module will replace missing values with 0 and then apply Label Encoder class to convert non-numeric values to numeric values and then shuffle and normalize dataset values
3. Features Selection: using this module will apply PCA algorithm on processed features to select relevant features from dataset and then split dataset into train and test where application using 70% dataset for training and 30% for testing
4. Run Existing Naive Bayes: 70% processed training data will be input to Naïve Bayes algorithm to train a model and this model will be applied on 30% test data to calculate prediction accuracy
5. Run Propose XGBoost: 70% processed training data will be input to Propose XGBOOST algorithm to train a model and this model will be applied on 30% test data to calculate prediction accuracy
6. Comparison Graph: using this module will plot comparison graph between existing and propose model
7. Predict Fraud from Test Data: using this module will upload test data and then apply XGBOOST algorithm to predict weather test data contains REAL or FRAUD transaction.

SCREEN SHOTS

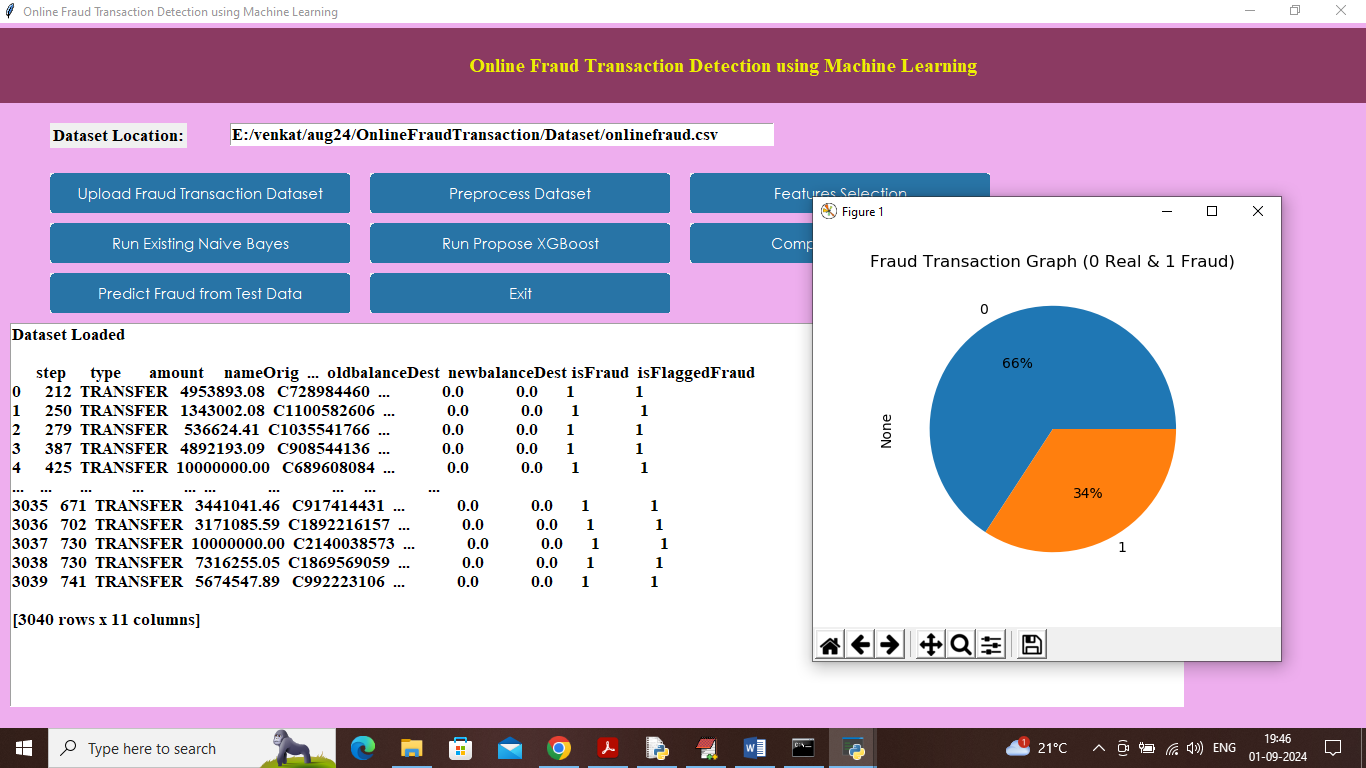
To run project double click on ‘run.bat’ file to get below screen



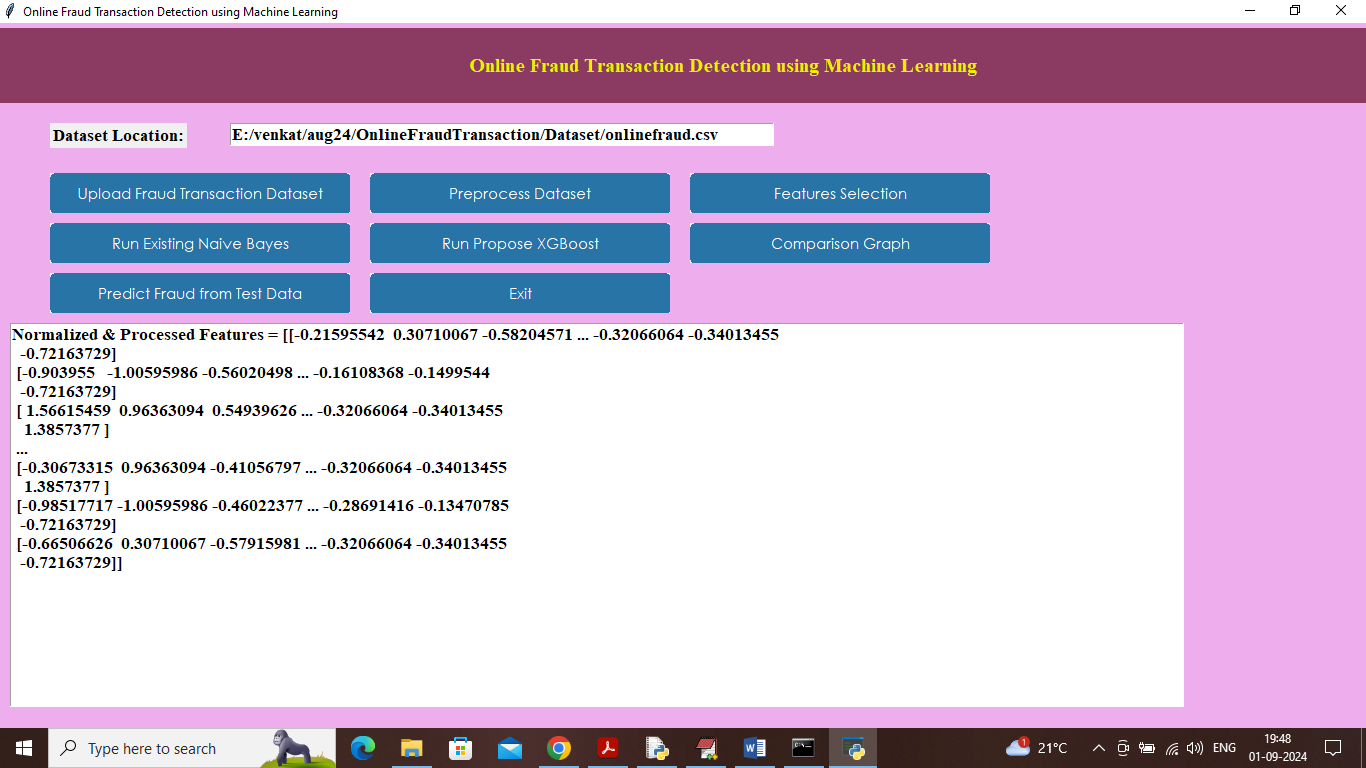
In above screen click on ‘Upload Fraud Transaction Dataset’ button to upload dataset and get below page



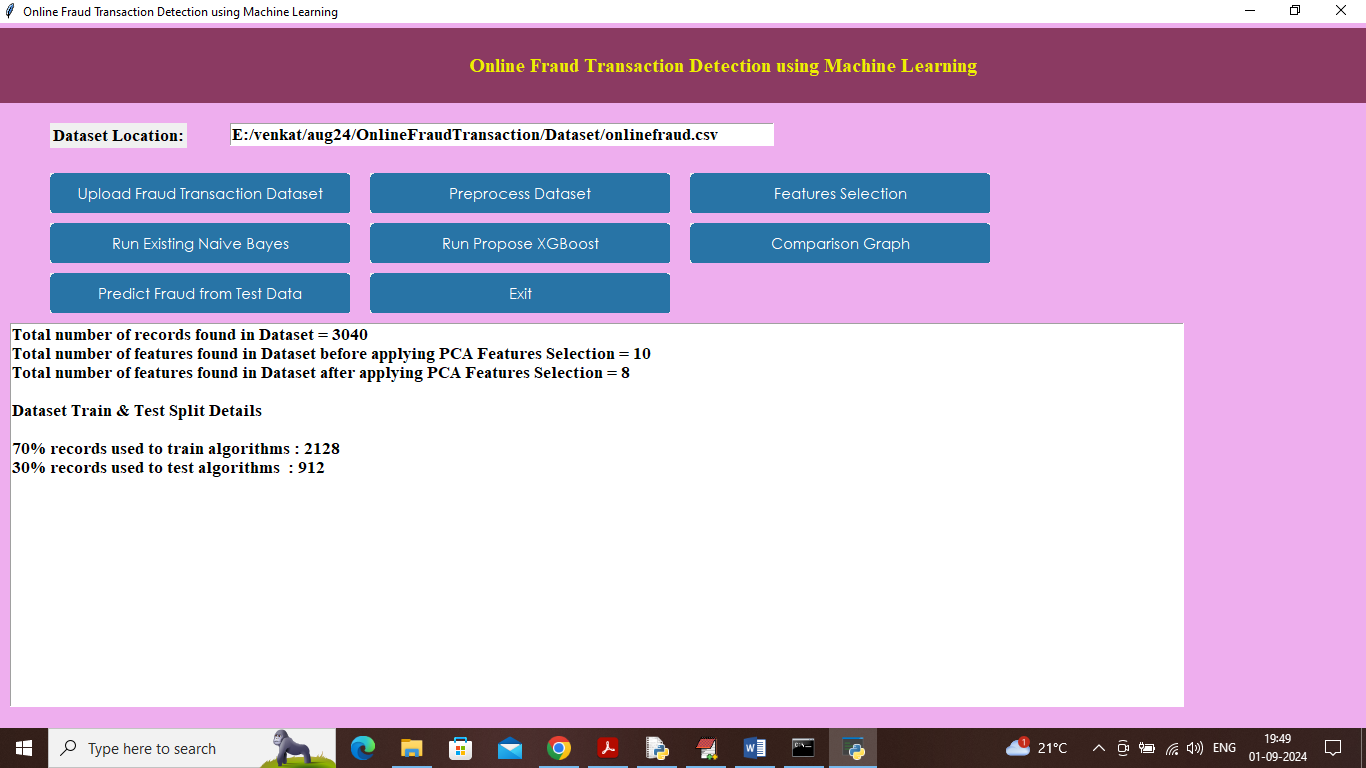
In above screen selecting and uploading dataset file and then click on ‘Open’ button to load dataset and get below page



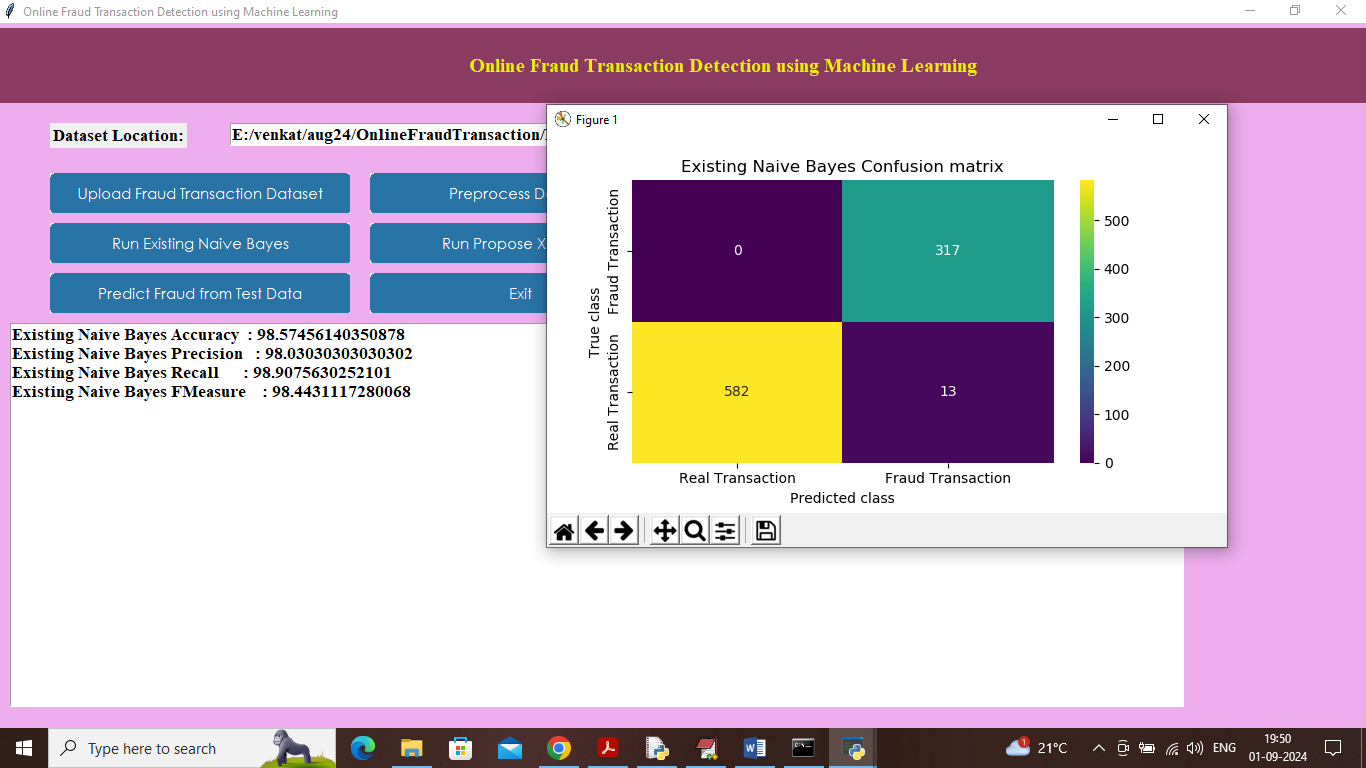
In above screen dataset loaded and in text area can see dataset values and in graph ‘0’ represents Real Transaction and 1 represents ‘Fraud’ transaction and then in graph can see percentage of Real and Fraud transaction available in dataset. Now close above graph and then click on ‘Pre-process Dataset’ button to clean and normalize dataset and get below page



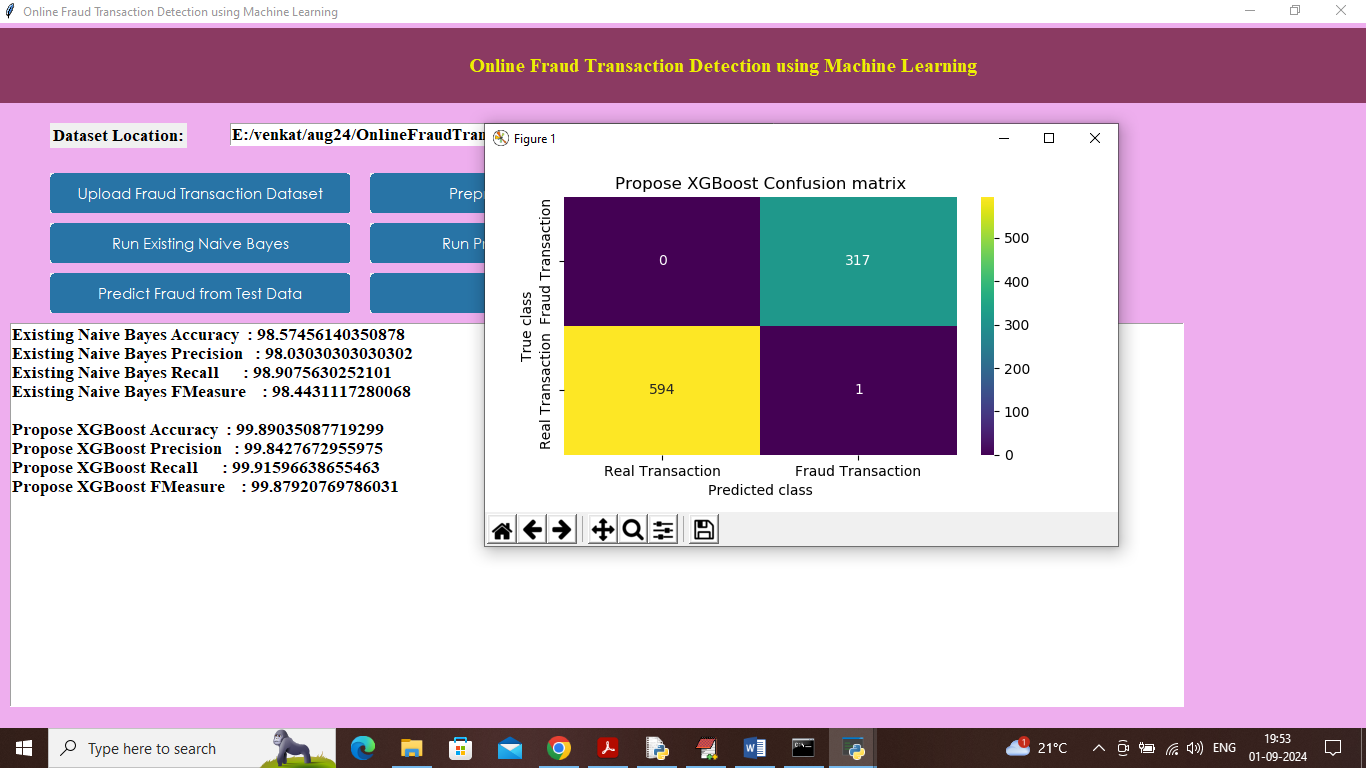
In above screen dataset processed and can see normalize features values and now click on ‘Features Selection’ button to select relevant features from the dataset and get below page



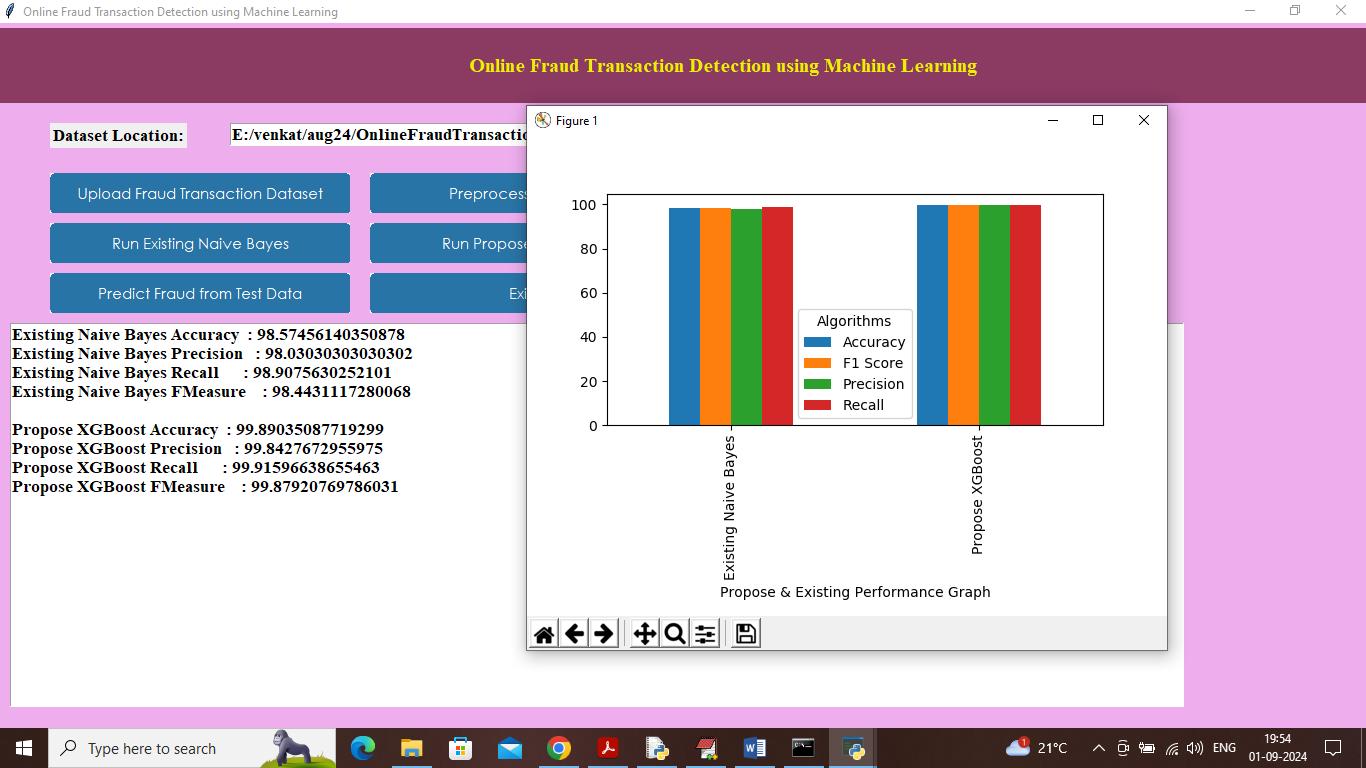
In above screen first 3 lines can see total features exists in dataset before applying PCA and then can see PCA selected 8 features out of 10 features and then can see training and testing dataset size. Now click on ‘Run Existing Naive Bayes’ button to train existing algorithm and get below output



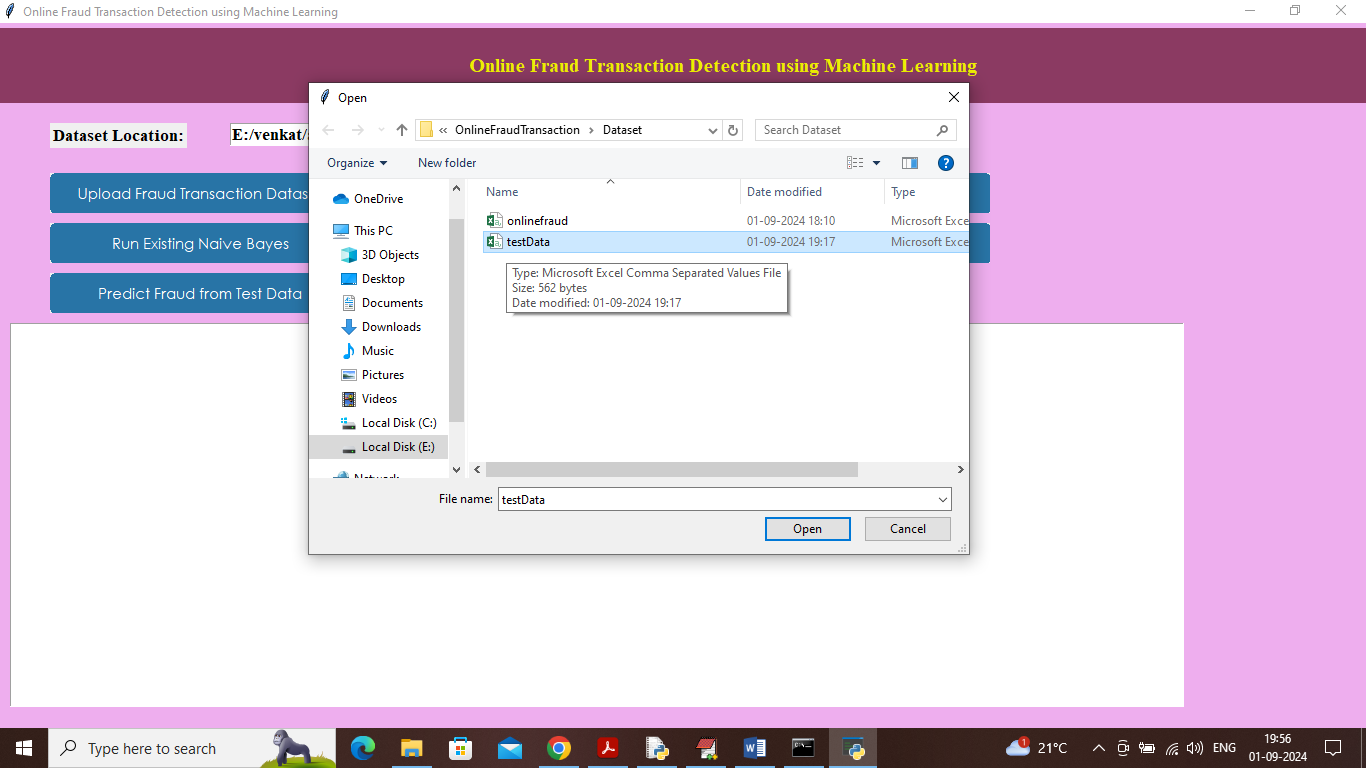
In above screen existing Naïve Bayes algorithm got 98% accuracy and can see other metrics like precision, recall and FSCORE. In confusion matrix graph x-axis represents ‘Predicted Labels’ and y-axis represents True Labels and then yellow and green boxes contains correct prediction count and all blue boxes contains incorrect prediction count which are very few. Now click on ‘Run Propose XGBoost’ button to train XGBOOST and get below page



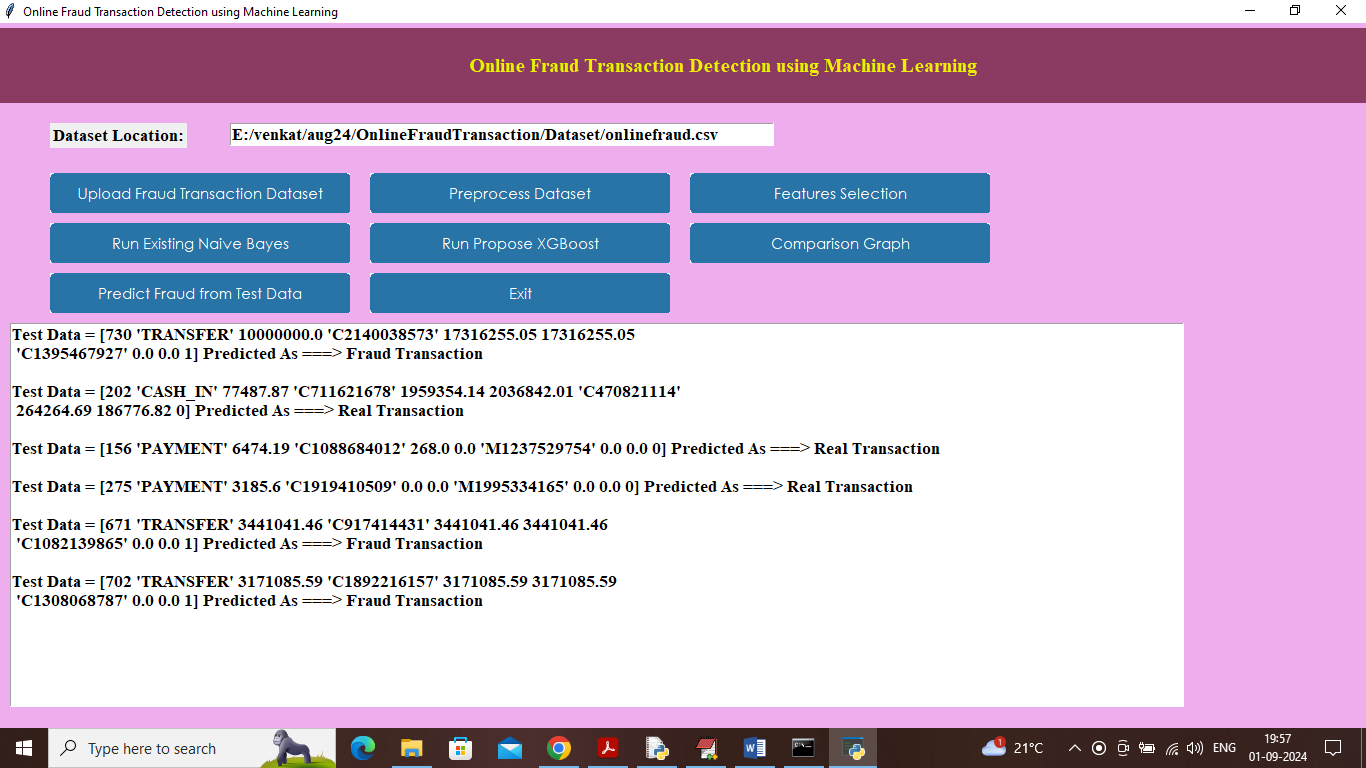
In above screen XGBOOST got 99% accuracy and can see other metrics also. Now click on ‘Comparison Graph’ button to get below graph



In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars and in all algorithms XGBOOST got high accuracy. Now close above graph and then click on ‘Predict Fraud from Test Data’ button to upload test data and get below prediction



In above screen selecting and uploading test data file and then click on ‘Open’ button to load dataset and get below output



In above screen in square bracket can see test data values and then after =🡺 symbol can see predicted values as ‘Real or Fraud’ transaction