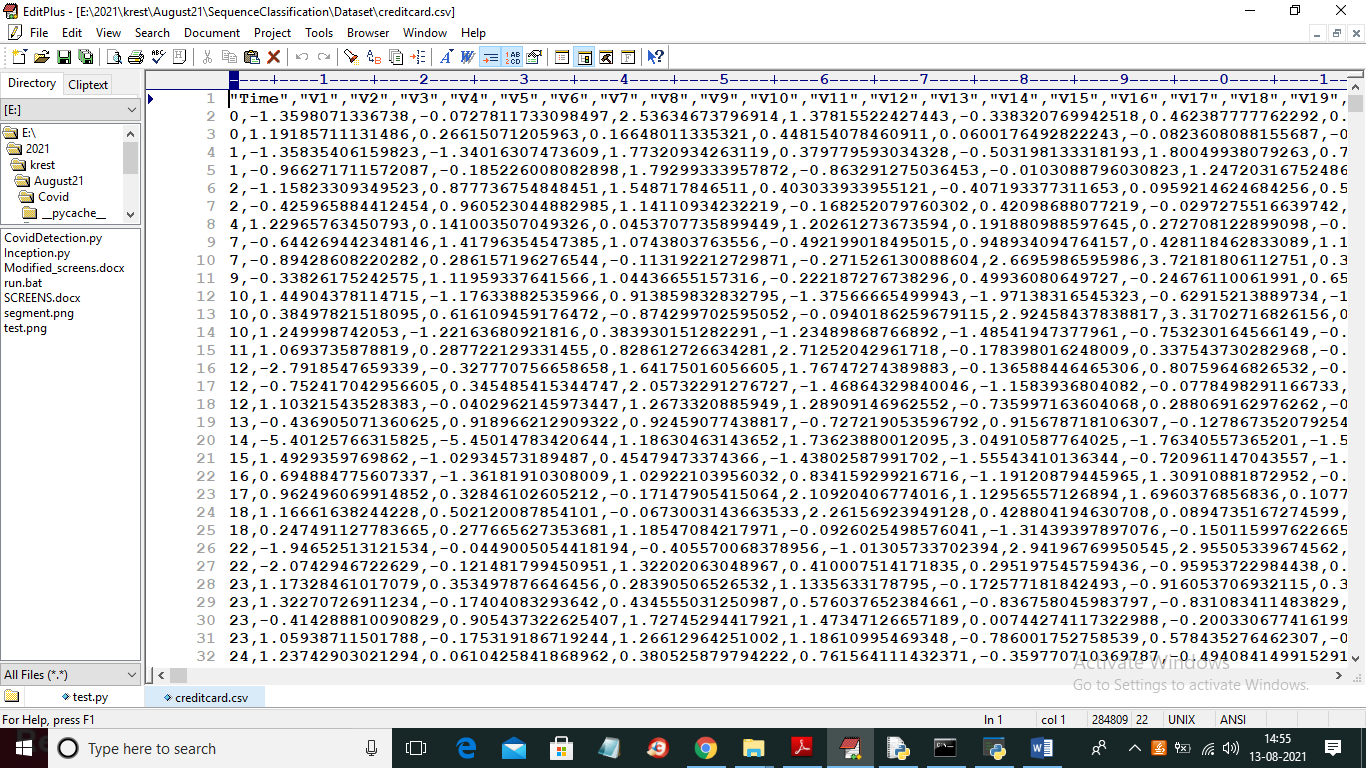
Sequence classification for credit-card fraud detection

In this paper author is using LSTM (long short term memory) neural network algorithm to detect fraud from credit card dataset by making sequence patterns. All existing algorithms will not consider time series data to detect fraud as over time fraudulent users may use different behaviour or technique to make fraud and LSTM is the only algorithm which make data as sequence of patterns and from this patterns he will predict fraud or non-fraud transaction. LSTM will trained itself using time series data and can able to predict fraud from new or old transaction.

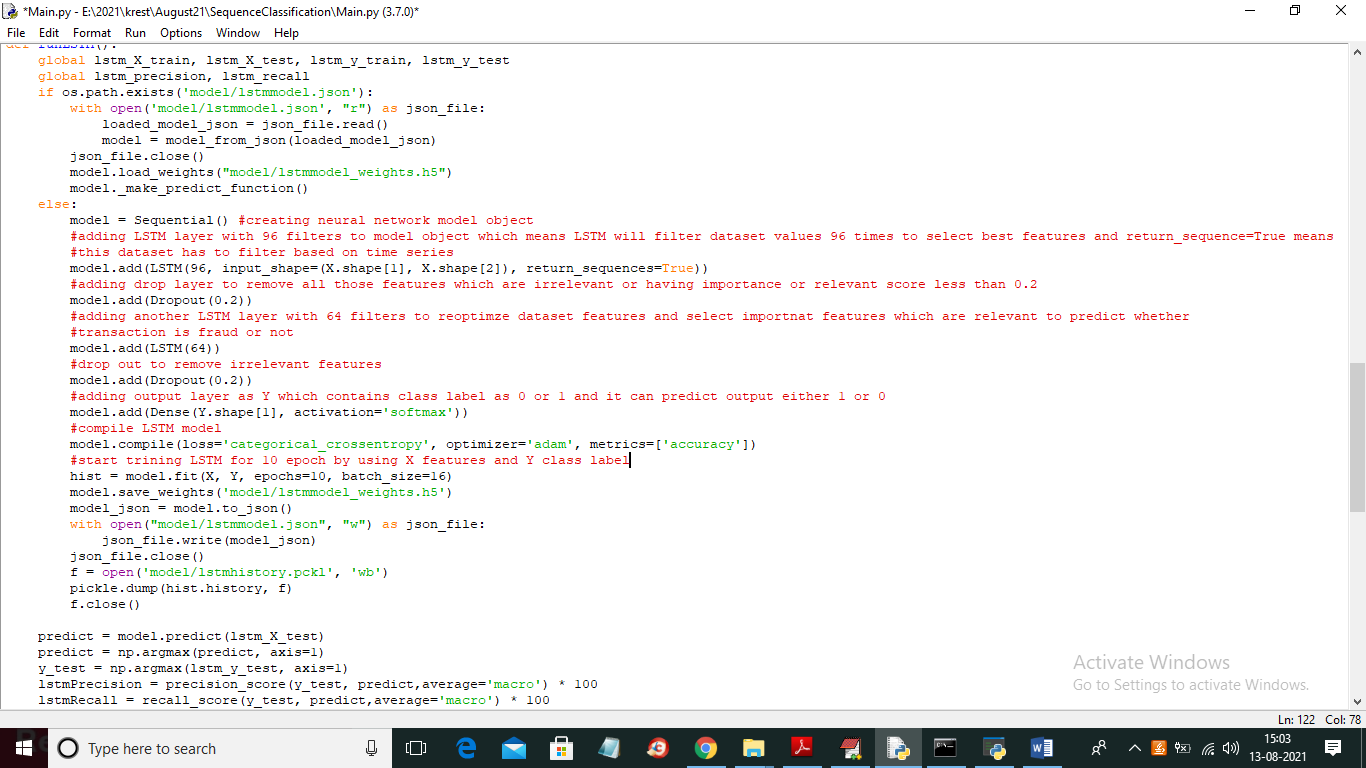
In propose paper author is comparing performance of LSTM sequence classifier with static (no sequence) Random Forest classifier.

In propose paper author has used Face 2 Face and ECOM transaction dataset and then divide that dataset into BASE, TDELTA and AGG. BASE will contains all raw features from dataset accept time information and TDELTA and AGG also take same features with time information. This dataset author has not publish on internet so I am using available credit card fraud dataset from KAGGLE website and I am using this dataset with all features and time information. Below is the dataset screen shots used to train LSTM and Random Forest Classifier.



In above dataset First row contains dataset column names and remaining rows contains anonymised transaction values and in last column each row is marked with class label as 0 or 1 where 0 means that transaction record is normal and 1 means fraud.

Below screen shot showing LSTM implementation with sequence pattern

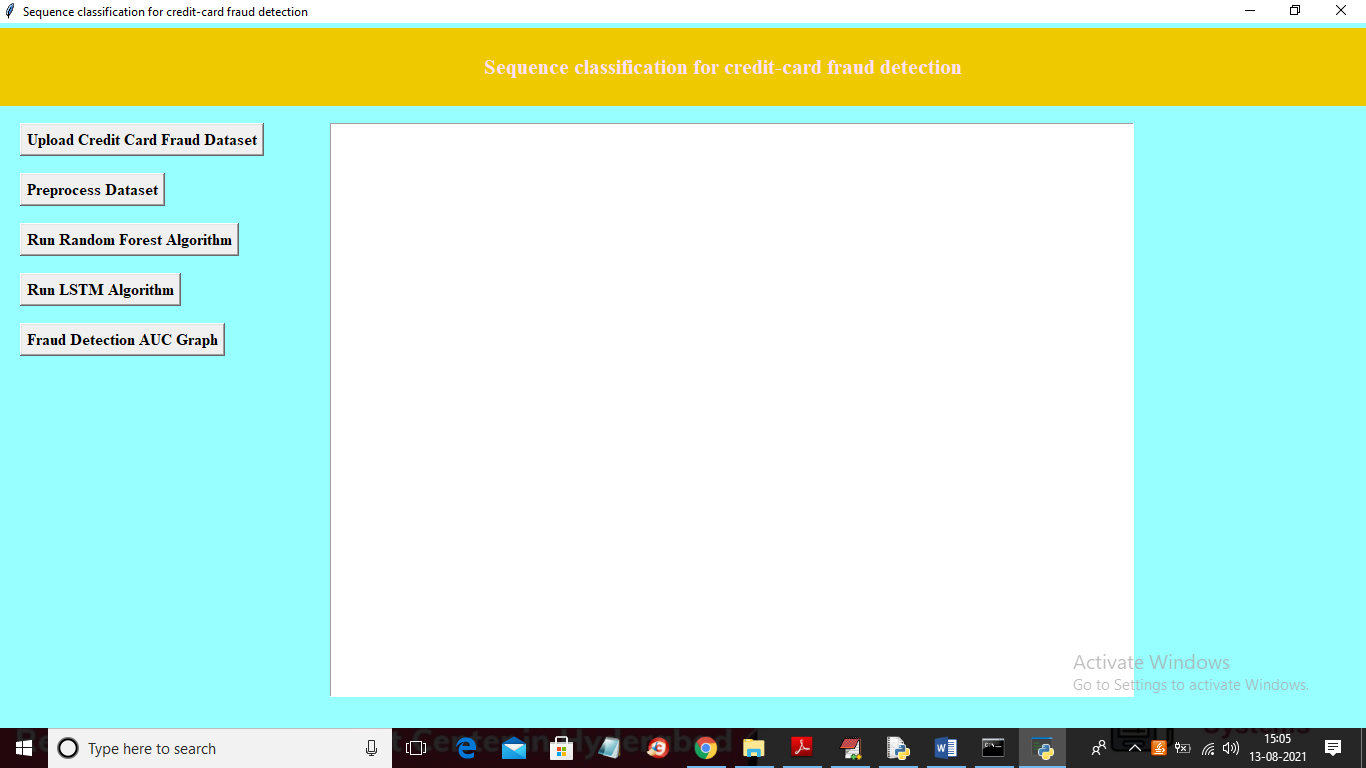


In above screen read red colour comments to know about LSTM implementation

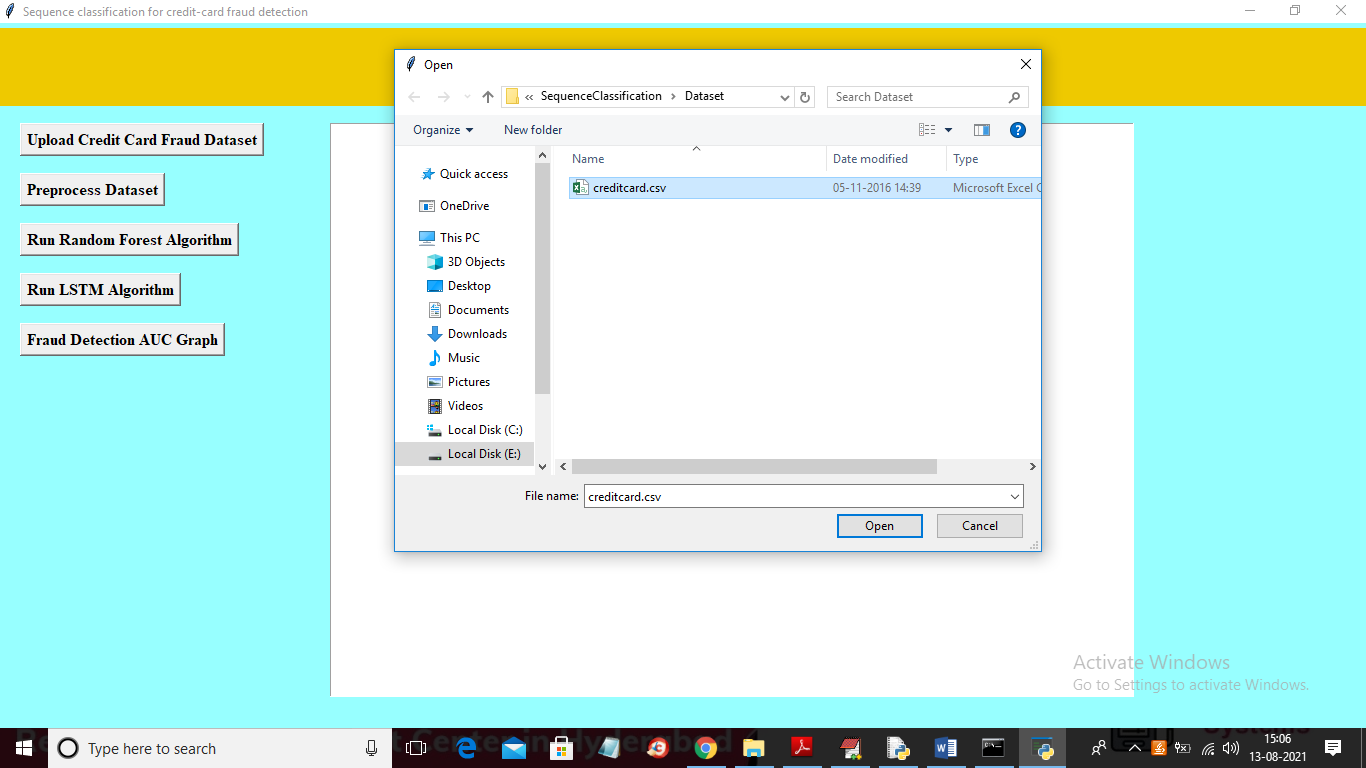
In propose paper author has divided dataset into base, delta and AGG and generate so many graphs but available dataset don’t have that many features so we can generate one graph

SCREEN SHOTS

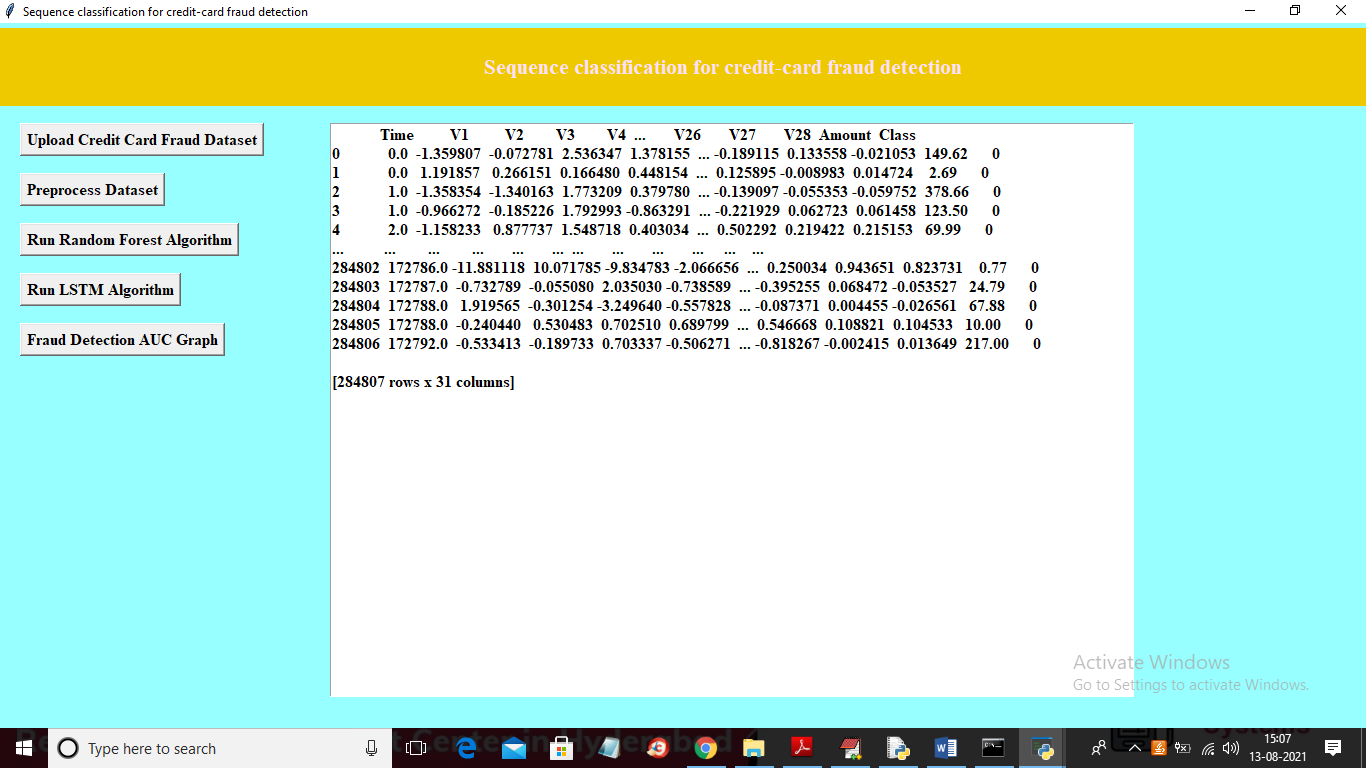
To run project double click on ‘run.bat’ file to get below screen



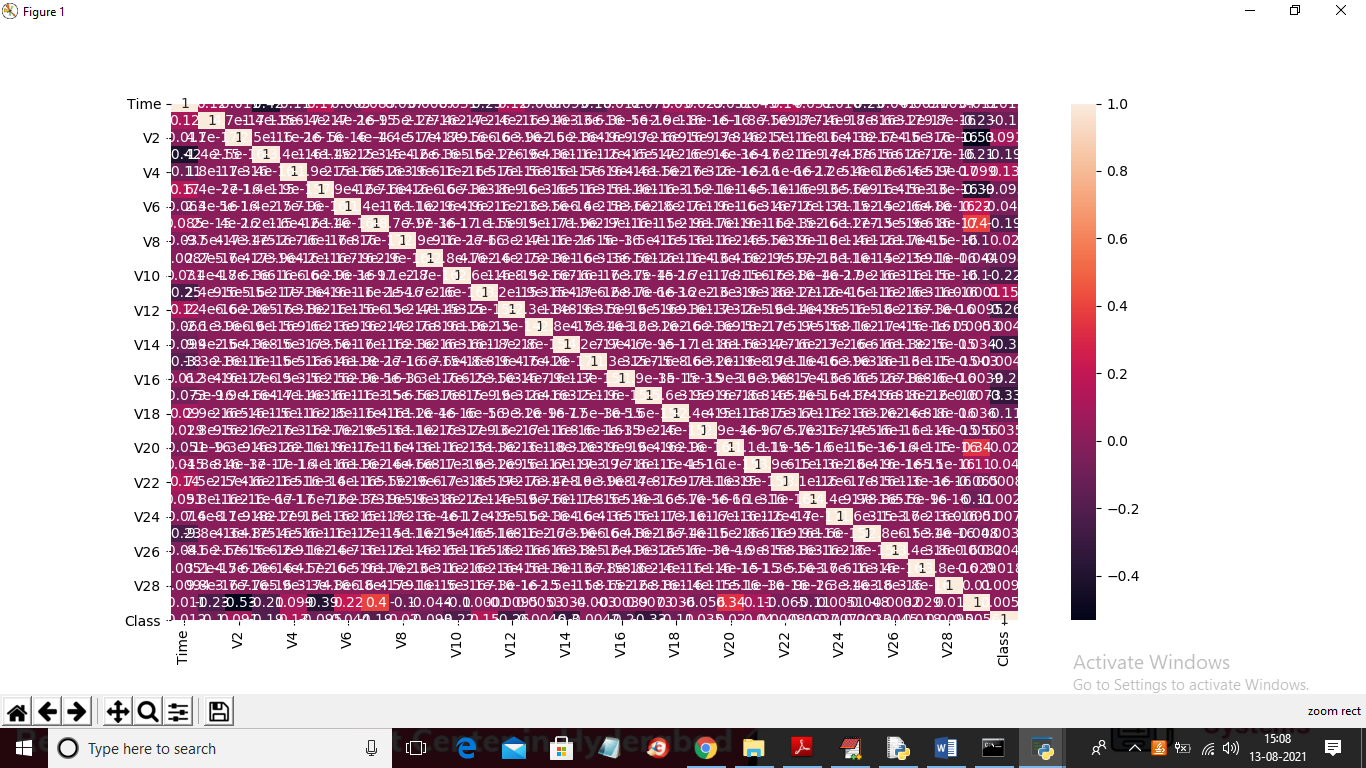
In above screen click on ‘Upload Credit Card Fraud Dataset’ button to upload dataset and to get below screen



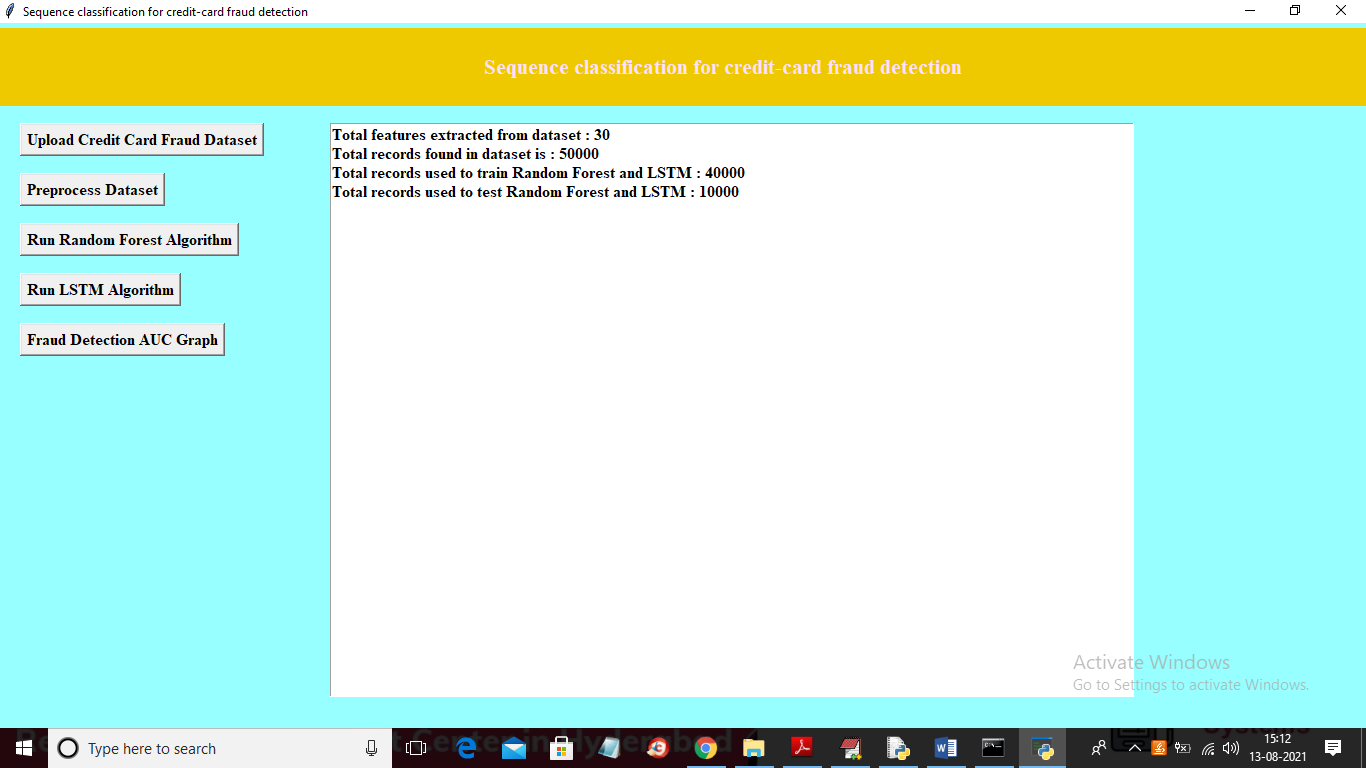
In above screen selecting and uploading ‘creditcard.csv’ file and then click on ‘Open’ button to load dataset and to get below screen



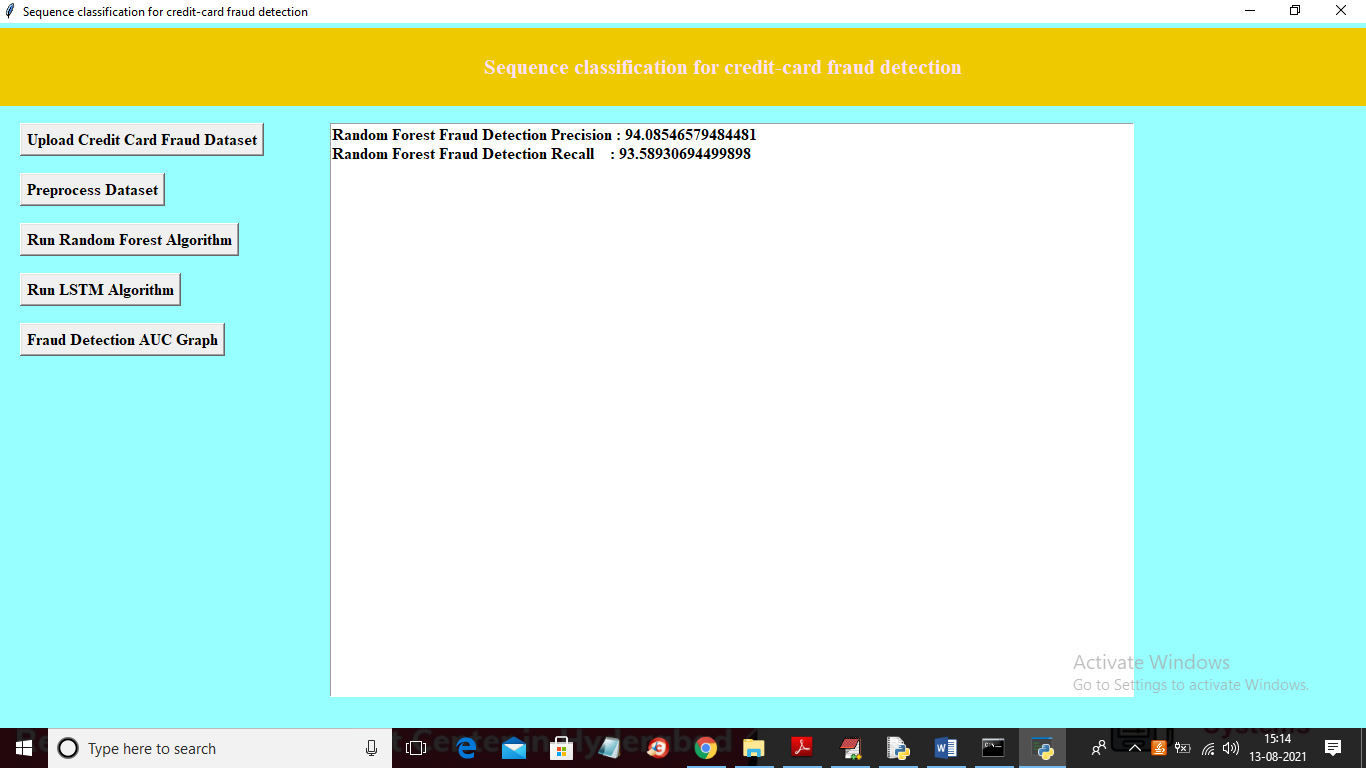
In above screen dataset loaded and in last column we can see values as 0 and similarly dataset will have 1 also and will get below graph which display features importance values



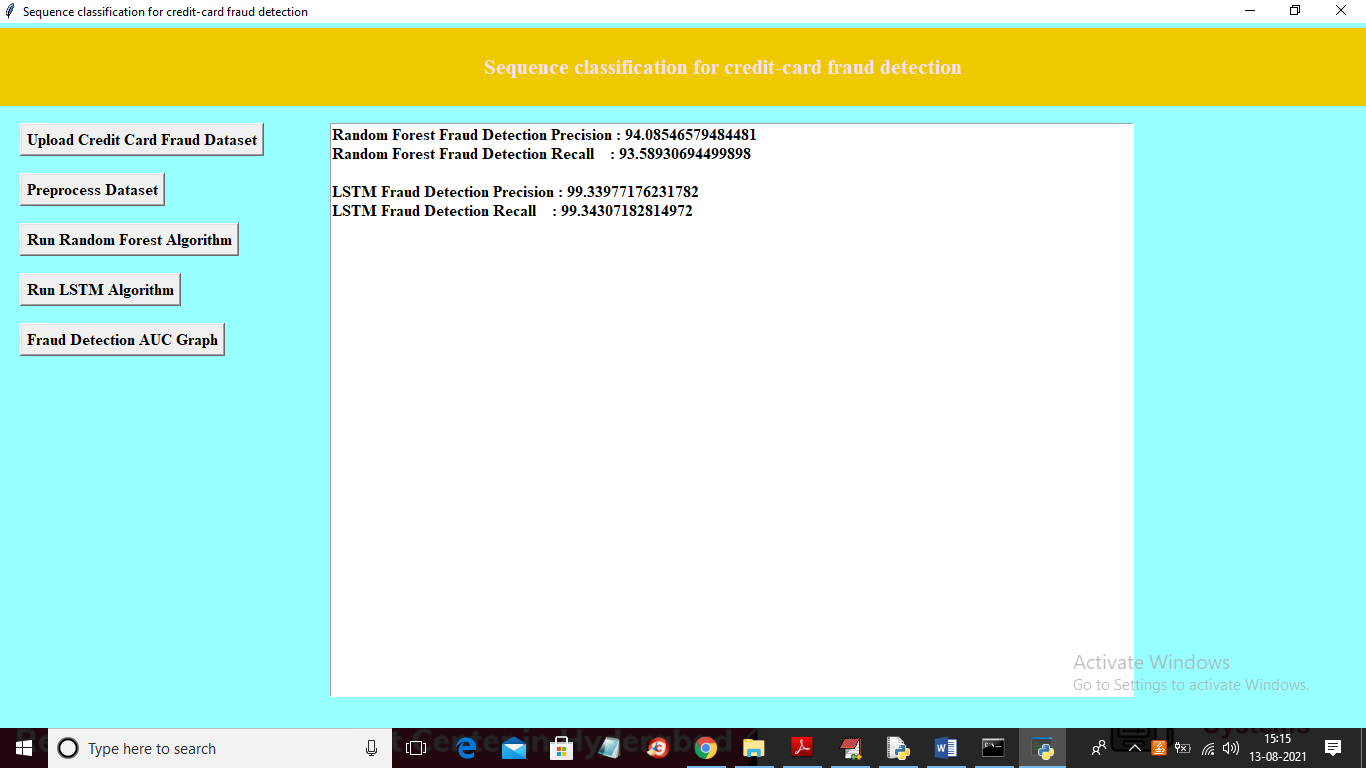
In above graph x and y-axis represents dataset features names and in boxes we can see features importance values and then features with high value will be consider as important features and LSTM will use only those features to train itself. Now close above graph and then click on ‘Preprocess Dataset’ button to remove missing values and then split dataset into train and test where random forest and LSTM will used train data to train themselves and test data to calculate their prediction performance as precision and recall.



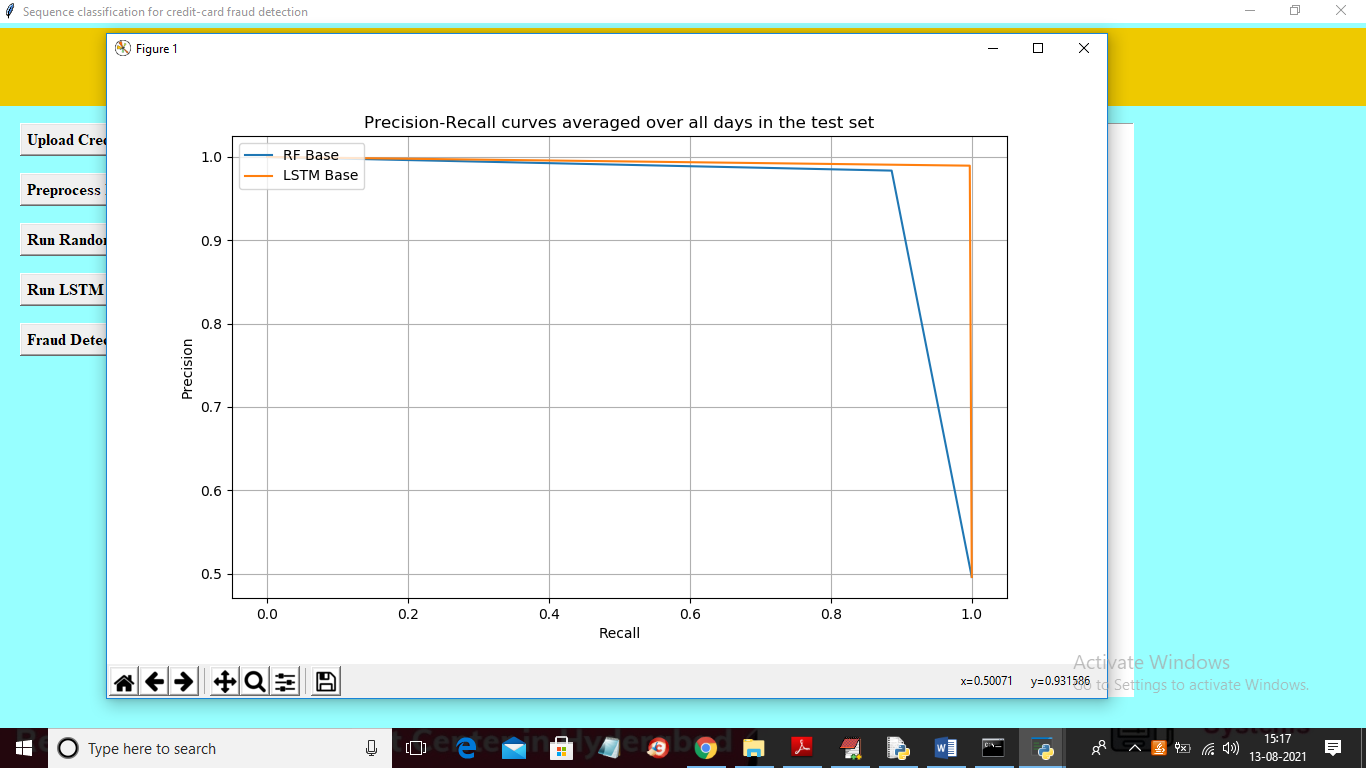
In above screen we can see dataset is processed and we got 30 features in the dataset and dataset contains 50000 records and application using 40000 records for training and 10000 for testing. Now train and test data is ready and now click on ‘Run Random Forest Algorithm’ button to train Random Forest



In above screen Random Forest training is completed and we got fraud detection precision as 94 and recall 93 and now click on ‘Run LSTM Algorithm’ button to train LSTM on same dataset and calculate its performance



In above screen with LSTM we got fraud detection precision and recall as 99% and due sequence pattern LSTM can able to predict fraud accurately so its performance will be little accurate compare to random forest. Now click on ‘Fraud Detection AUC Graph’ button to get below comparison of both algorithms



In above graph x-axis represents recall and y-axis represents precision and blue line represents random forest and orange line represents LSTM and in above graph we can see LSTM performance is close to 1 and random forest is less than 1. From above graph we can conclude that LSTM is better in fraud detection compare to Random Forest.