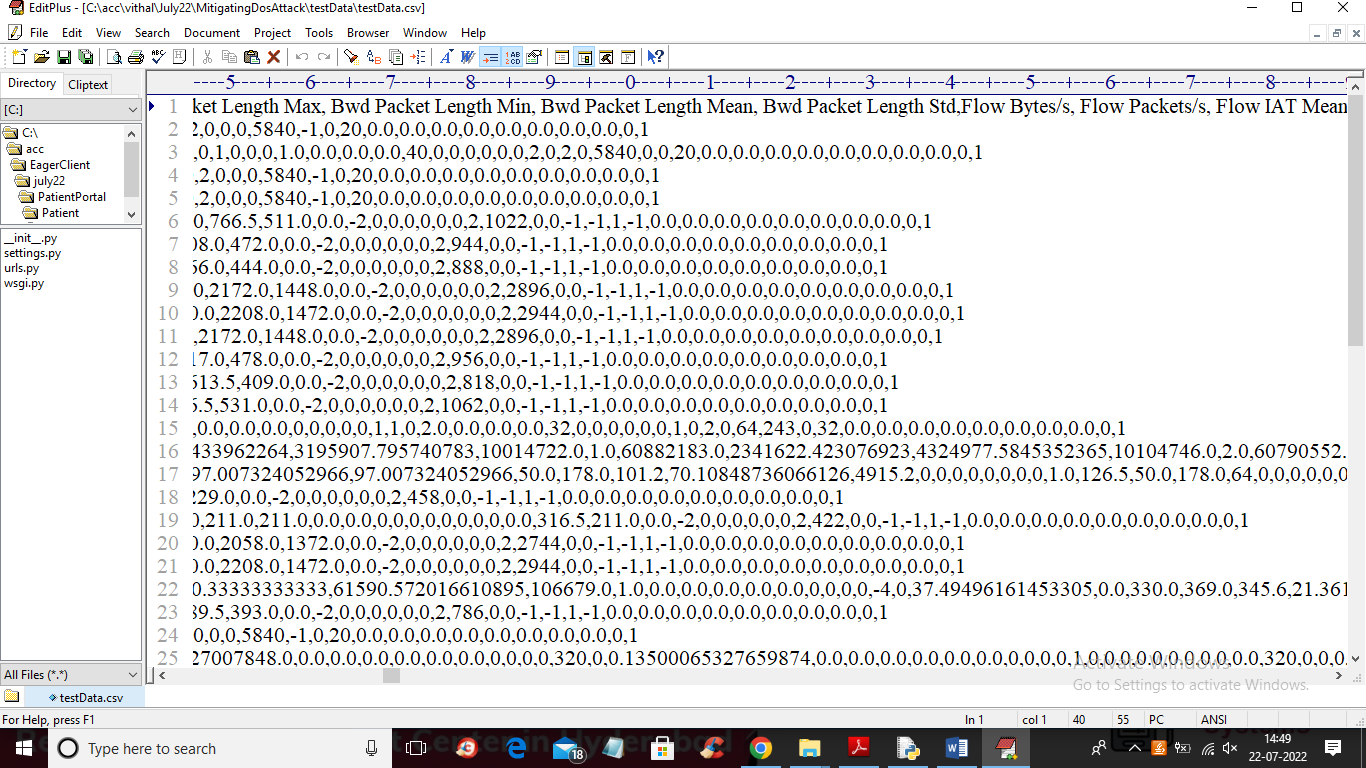
In this project we have used various machine learning algorithms such as SVM, Random Forest, XGBOOST, ADABOOST, KNN and Naïve Bayes and in all this algorithms Random Forest is giving best accuracy more than 95%.

To train this algorithm we have used CIC dataset which contains 10 different attacks of IOT environment and 1 BENIGN (normal) class.

To implement this project we have designed following modules

1. Upload DDOS Dataset: using this module we will upload dataset and then application will read data from all attacks files and then plot graph of different attacks. Dataset contains both numeric and non-numeric data but machine learning algorithms will accept only numeric data so by applying preprocessing technique we need to convert all non-numeric data to numeric.
2. Preprocess Dataset: using this module we will replace missing values with 0 and then apply Label Encoder technique which will assign numeric ID to each non-numeric data and then split dataset into TRAIN and TEST where application used 80% dataset for training and 20% for testing
3. Run Naive Bayes Algorithm: using this module we will input 80% train data to Naïve Bayes algorithm to train a model and this model will be applied on test data to calculate prediction accuracy.
4. Run Random Forest Algorithm: using this module we will input 80% train data to Random Forest algorithm to train a model and this model will be applied on test data to calculate prediction accuracy.
5. Run SVM Algorithm: using this module we will input 80% train data to SVM algorithm to train a model and this model will be applied on test data to calculate prediction accuracy.
6. Run XGBOOST Algorithm: using this module we will input 80% train data to XGBOOST algorithm to train a model and this model will be applied on test data to calculate prediction accuracy.
7. Run ADABOOST Algorithm: using this module we will input 80% train data to ADABOOST algorithm to train a model and this model will be applied on test data to calculate prediction accuracy.
8. Run KNN Algorithm: using this module we will input 80% train data to KNN algorithm to train a model and this model will be applied on test data to calculate prediction accuracy.
9. Comparison Graph: using this module we will display comparison table and graph of all algorithms
10. Predict Attack from Test Data: using this module we will upload test data and then machine learning models will predict attack from that test data.

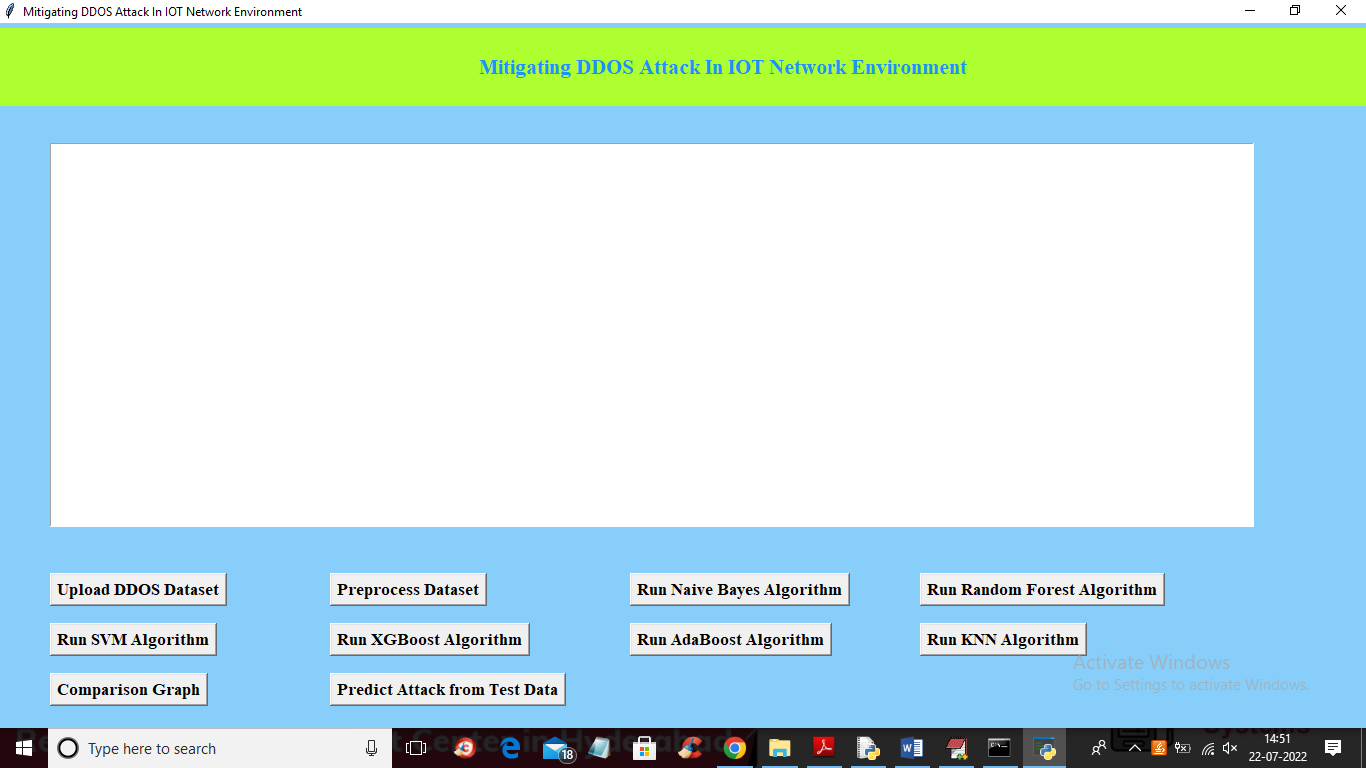
Test Data you can find inside test folder and this test data contains all features without any class label and this label will be predicted by machine learning model. Below is the test data screen



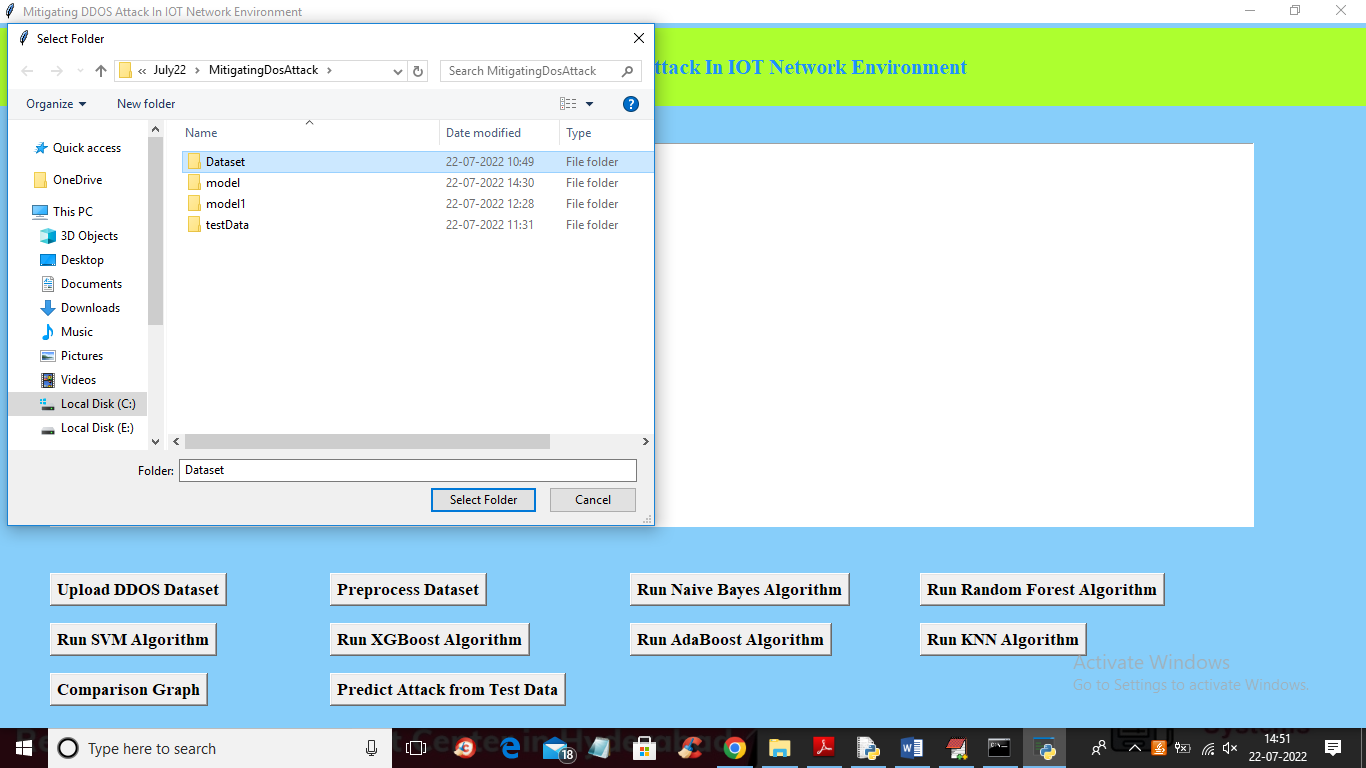
In above TEST DATA screen there is no class label or attack name and this will be predicted by ML model.

SCREEN SHOTS

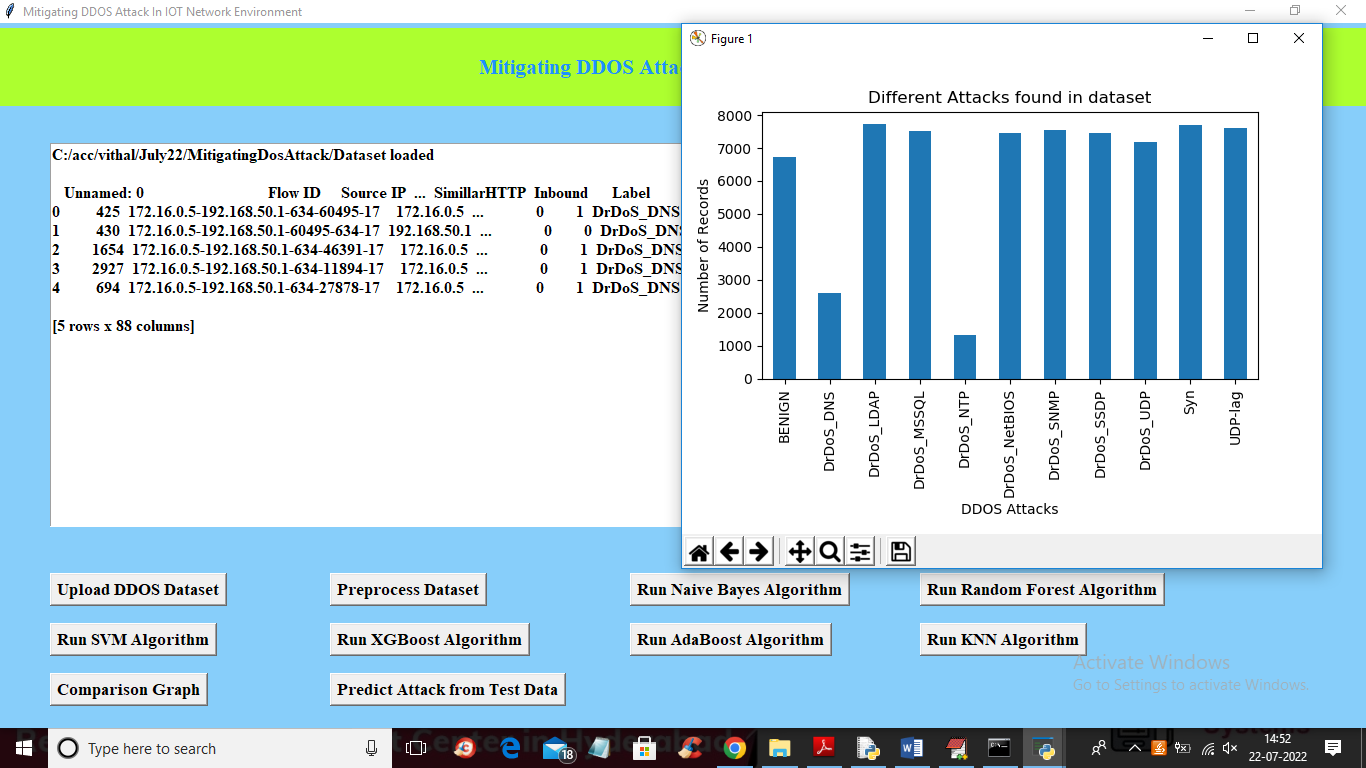
To run project double click on ‘run.bat’ file to get below screen



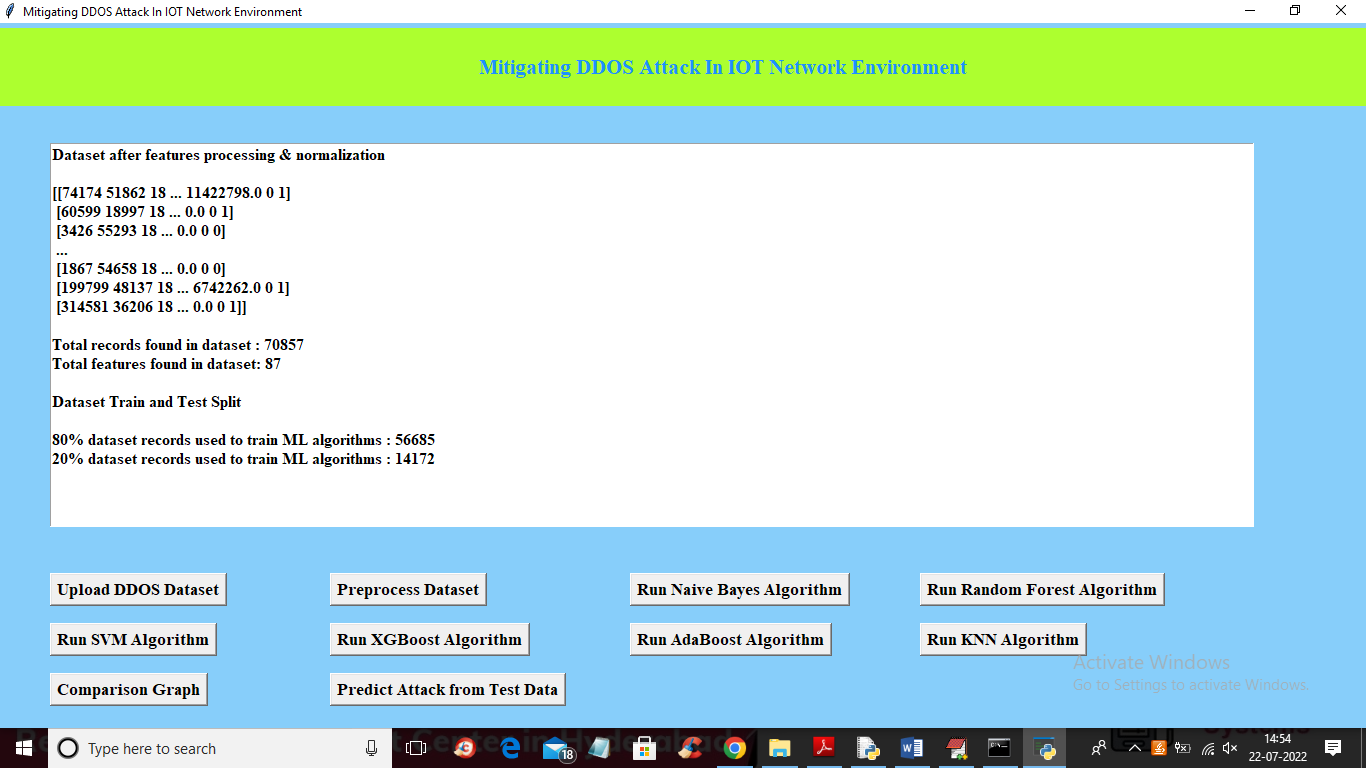
In above screen click on ‘Upload DDOS Dataset’ button tto upload dataset and get below output



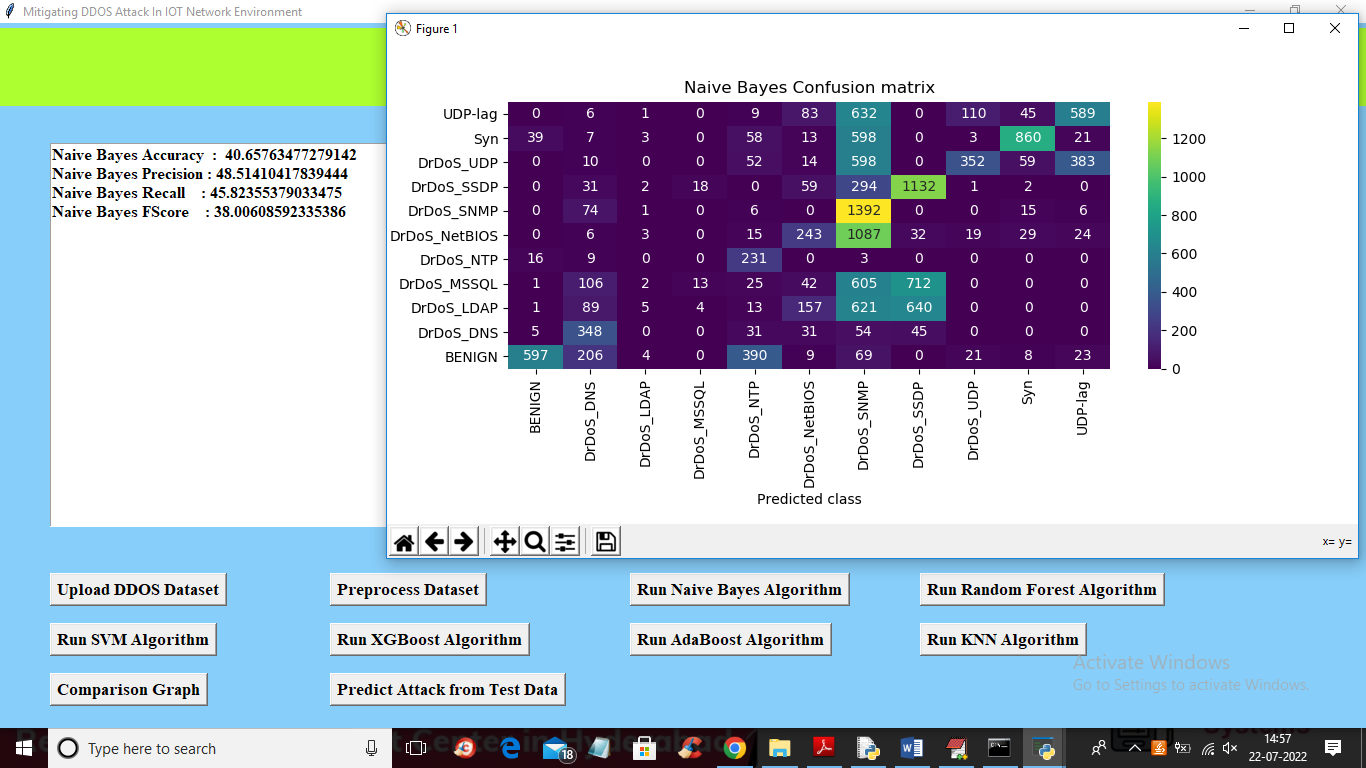
In above screen selecting and uploading ‘Dataset’ folder and then click on ‘Select Folder’ button to load dataset and get below output



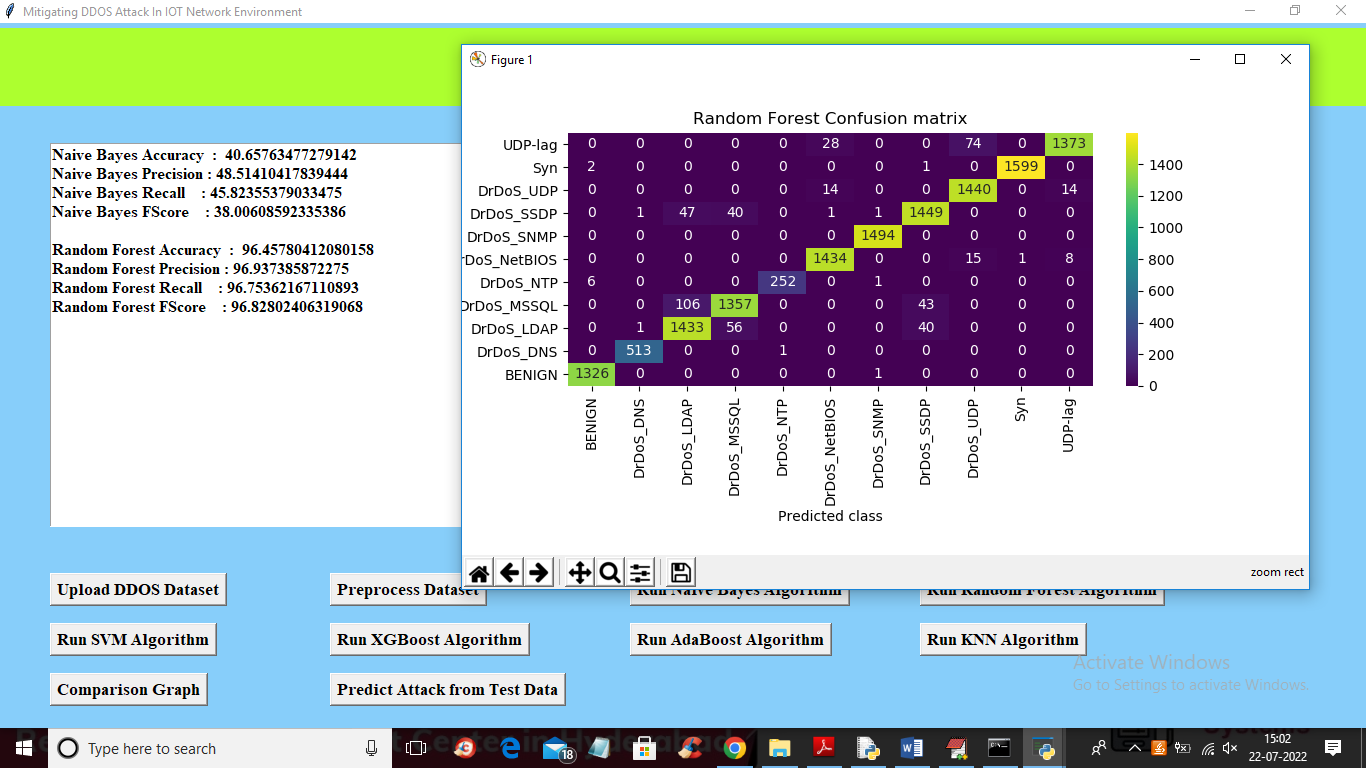
In above screen dataset loaded and we can see dataset contains both numeric and non-numeric data and in above graph x-axis represents attack names and y-axis represents count of those records. Now close above graph and then click on ‘Preprocess Dataset’ button to process dataset and get below screen



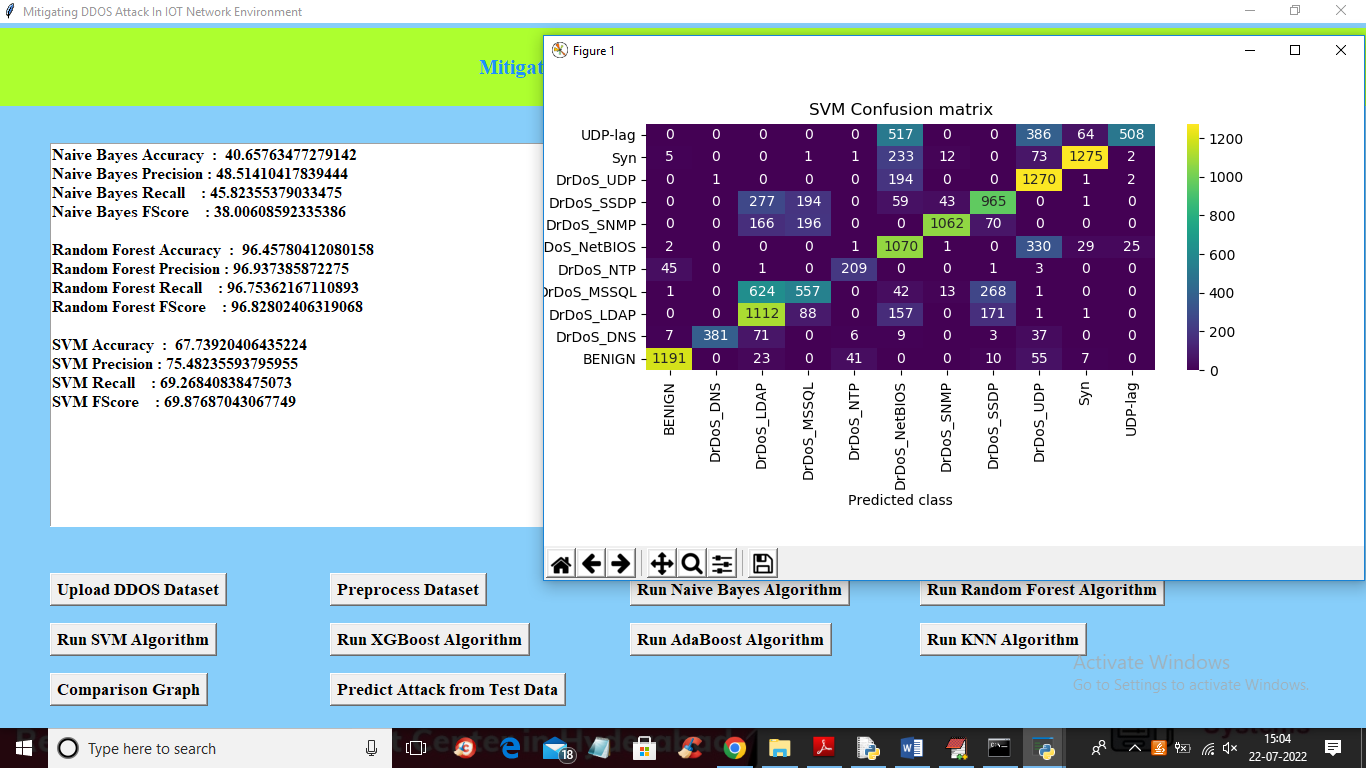
In above screen we can see all dataset values converted to numeric format and dataset contains more than 70000 records and each record contains 87 features and then we have split dataset into train and test and for training application using 56685 records for training and 14172 for testing. Now train and test data is ready and now click on ‘Run Naïve Bayes Algorithm’ button to train Naïve Bayes and get below output



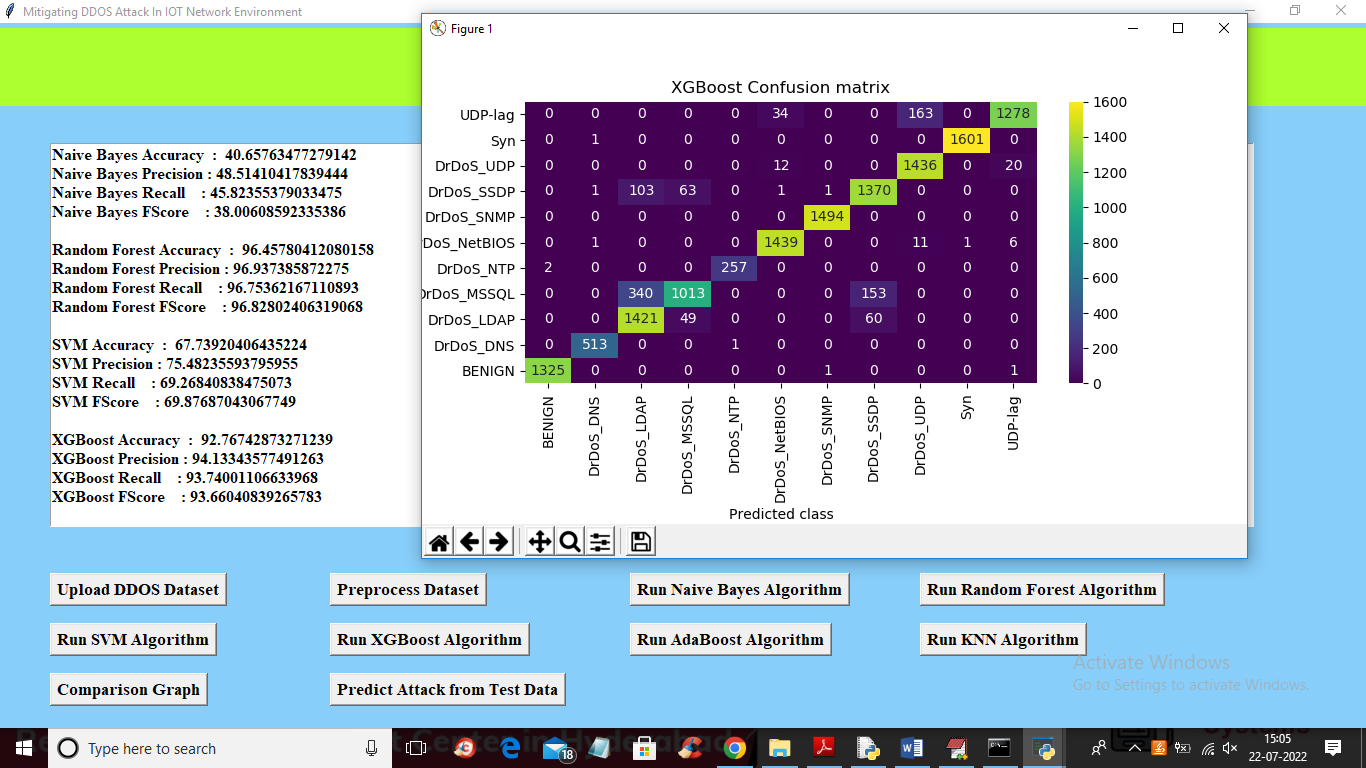
In above screen with Naïve Bayes we got 40% accuracy and in confusion matrix graph x-axis represents PREDICTED classes and y-axis represents TRUE classes and prediction count in same row and column names are the correct prediction and count in different row and column names are the incorrect prediction and we can see Naïve Bayes predicted so many wrong prediction and close above graph and then click on ‘Run Random Forest Algorithm’ button to get below output



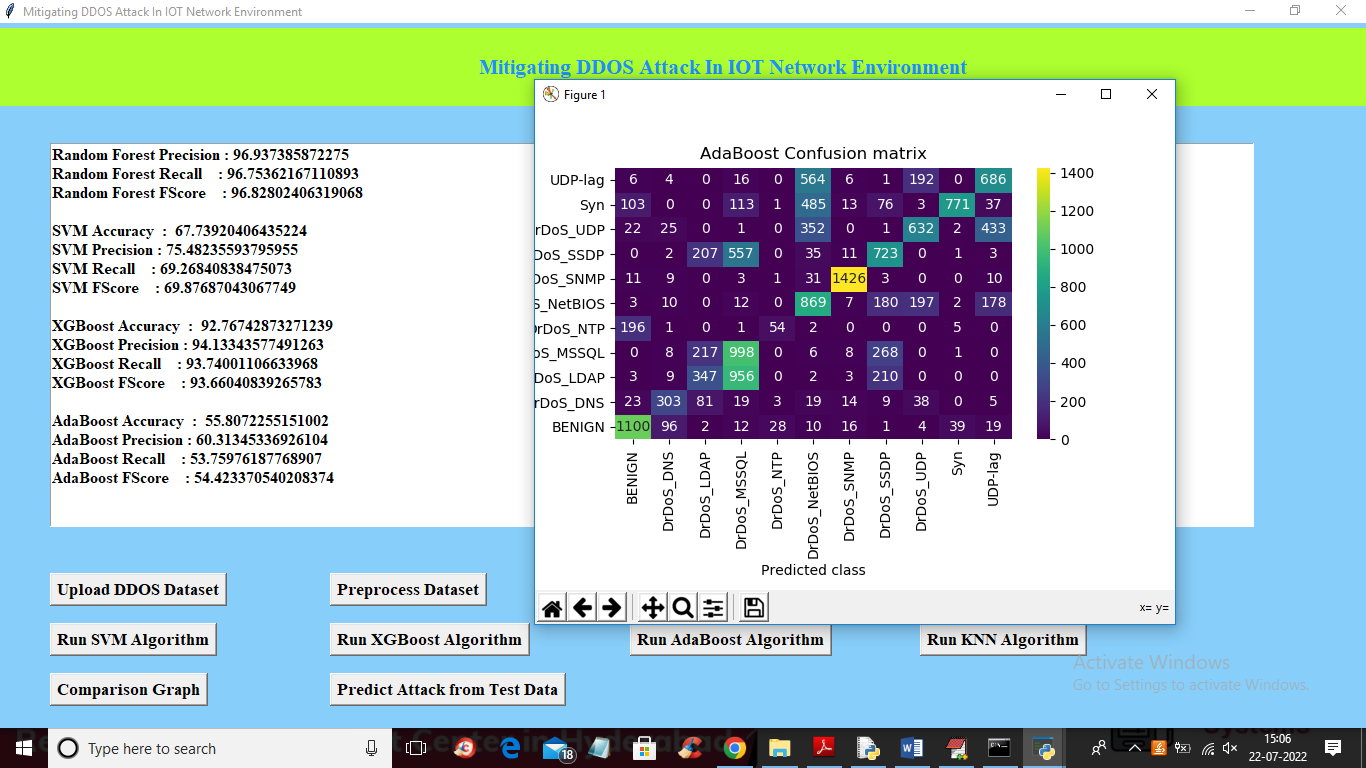
In above screen with Random Forest we got more than 96% accuracy and in graph also we can see lots of predictions are correct. Now close above graph and then click on ‘Run SVM Algorithm’ button to get below output



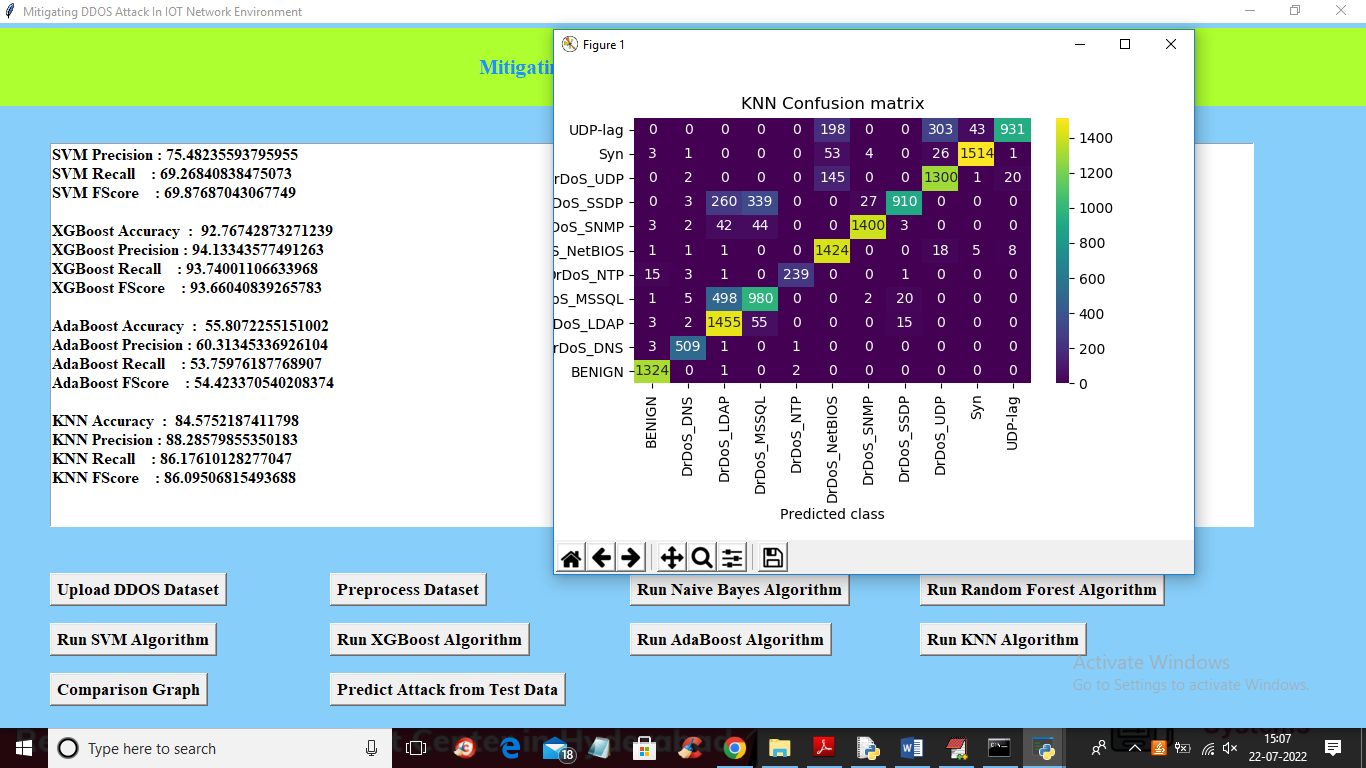
In above screen with SVM we got 67% accuracy and now close above graph and then click on ‘Run XGBOOST Algorithm’ button to get below output



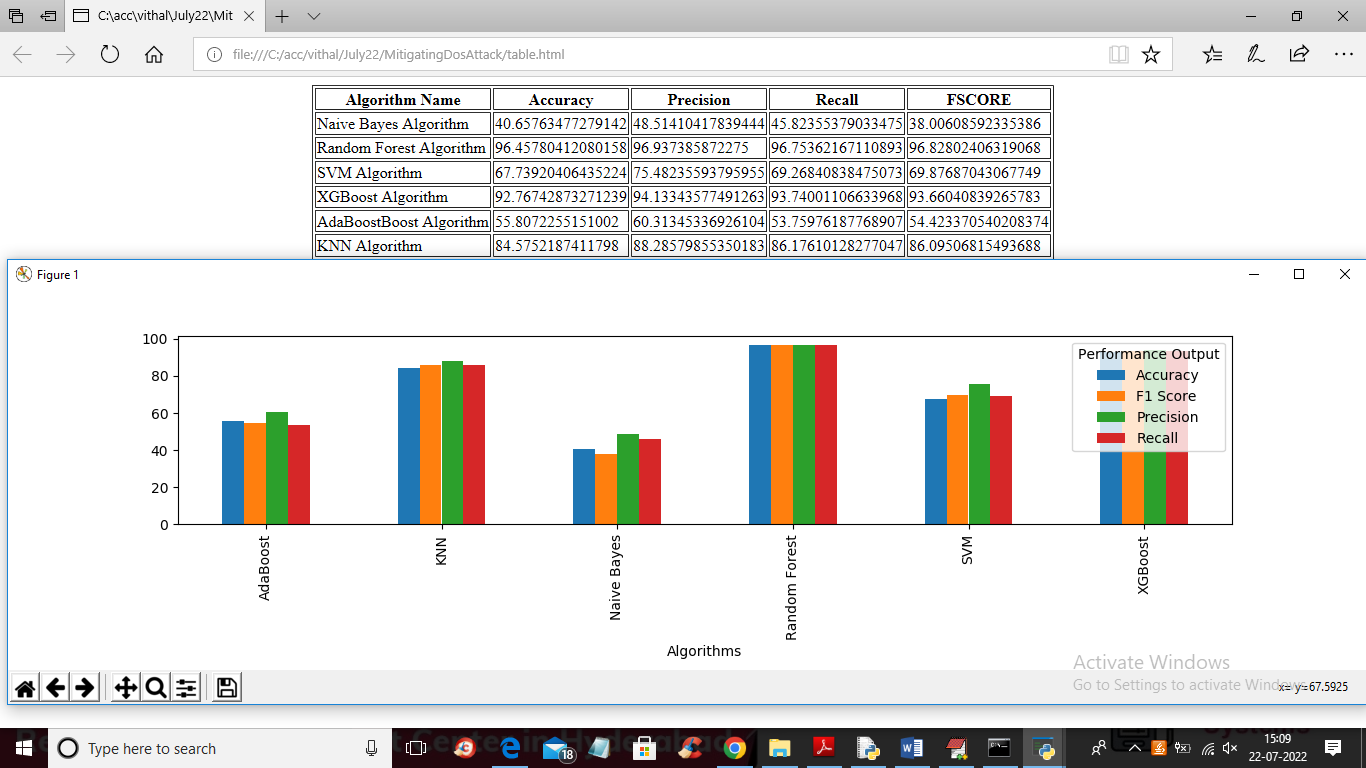
In above screen with XGBOOST we got 92% accuracy and now close above graph and then click on ‘Run ADA BOOST Algorithm’ button to get below output



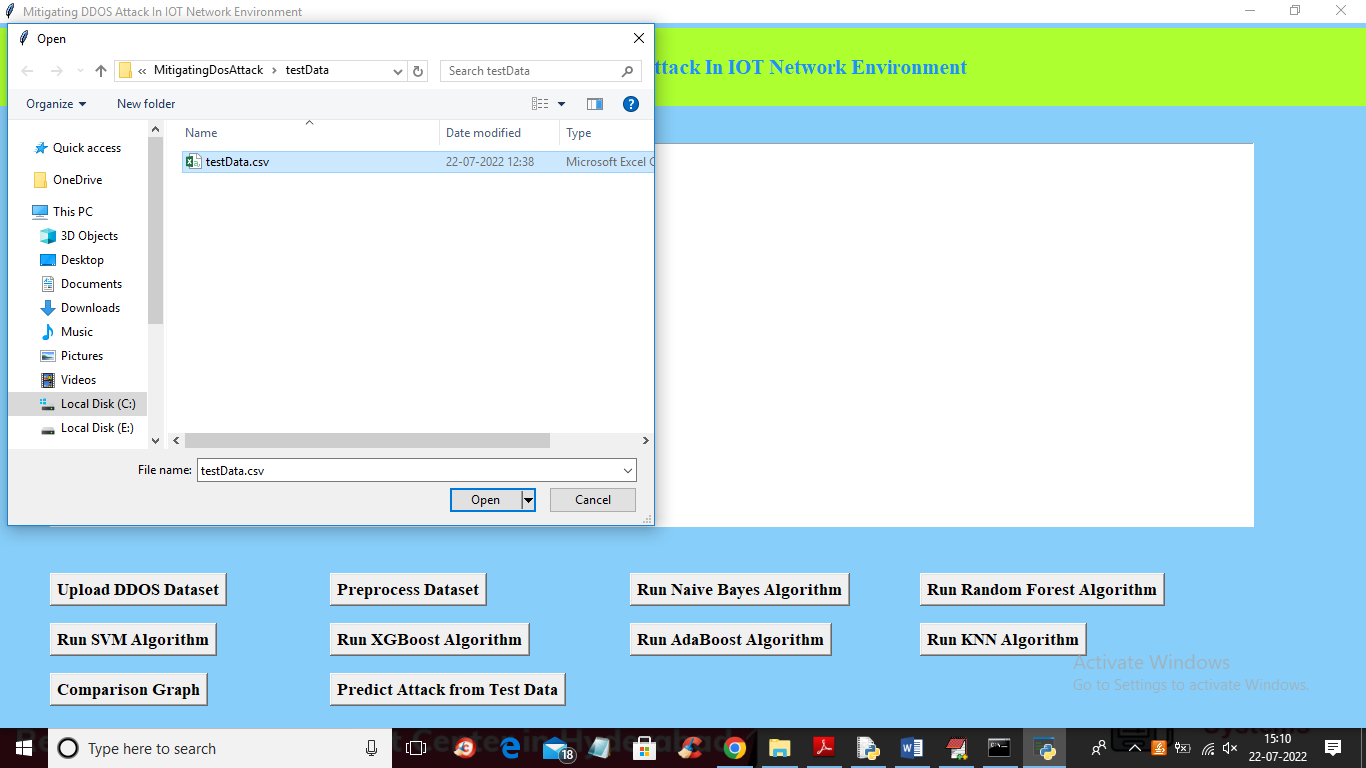
In above screen with ADABOOST we got 55% accuracy and now close above graph and then click on ‘Run KNN Algorithm’ button to get below output



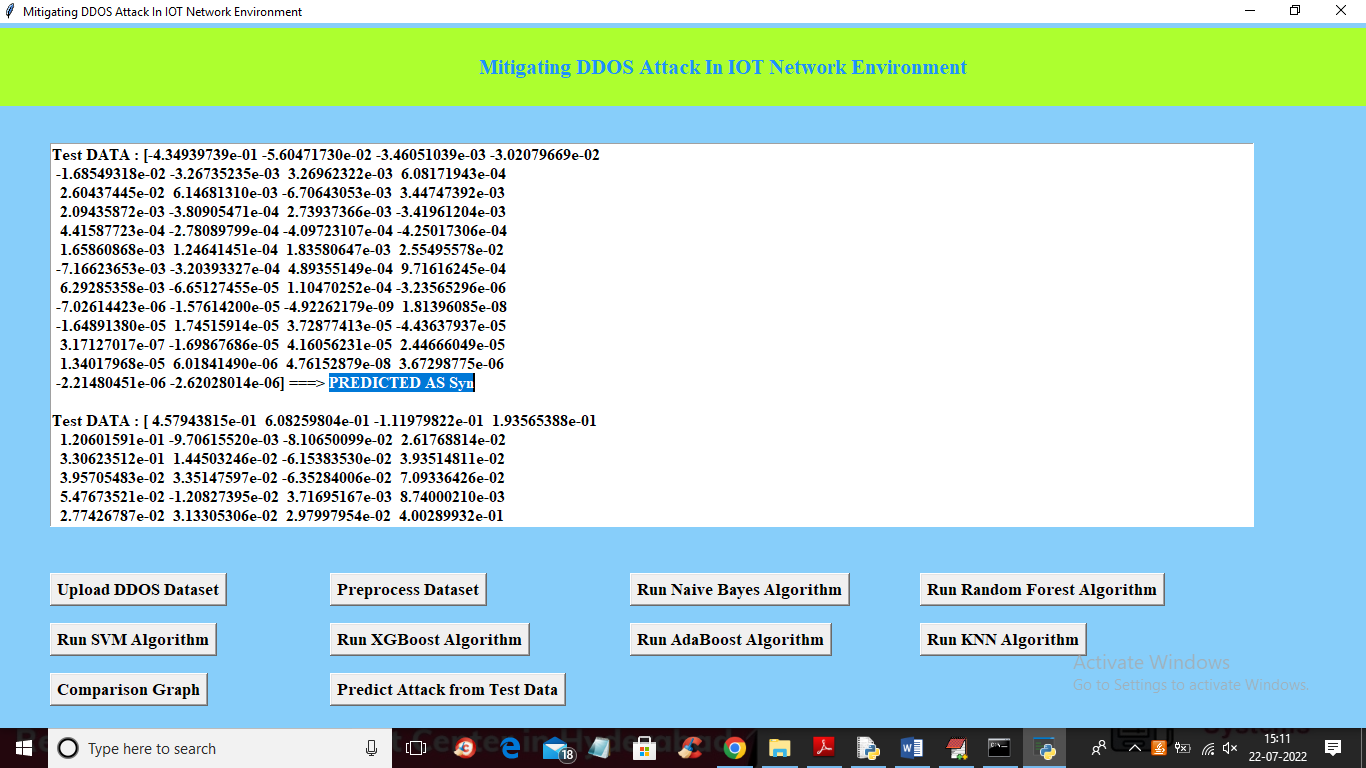
In above screen with KNN we got 84% accuracy and now close above graph and then click on ‘Comparison Graph’ button to get below graph



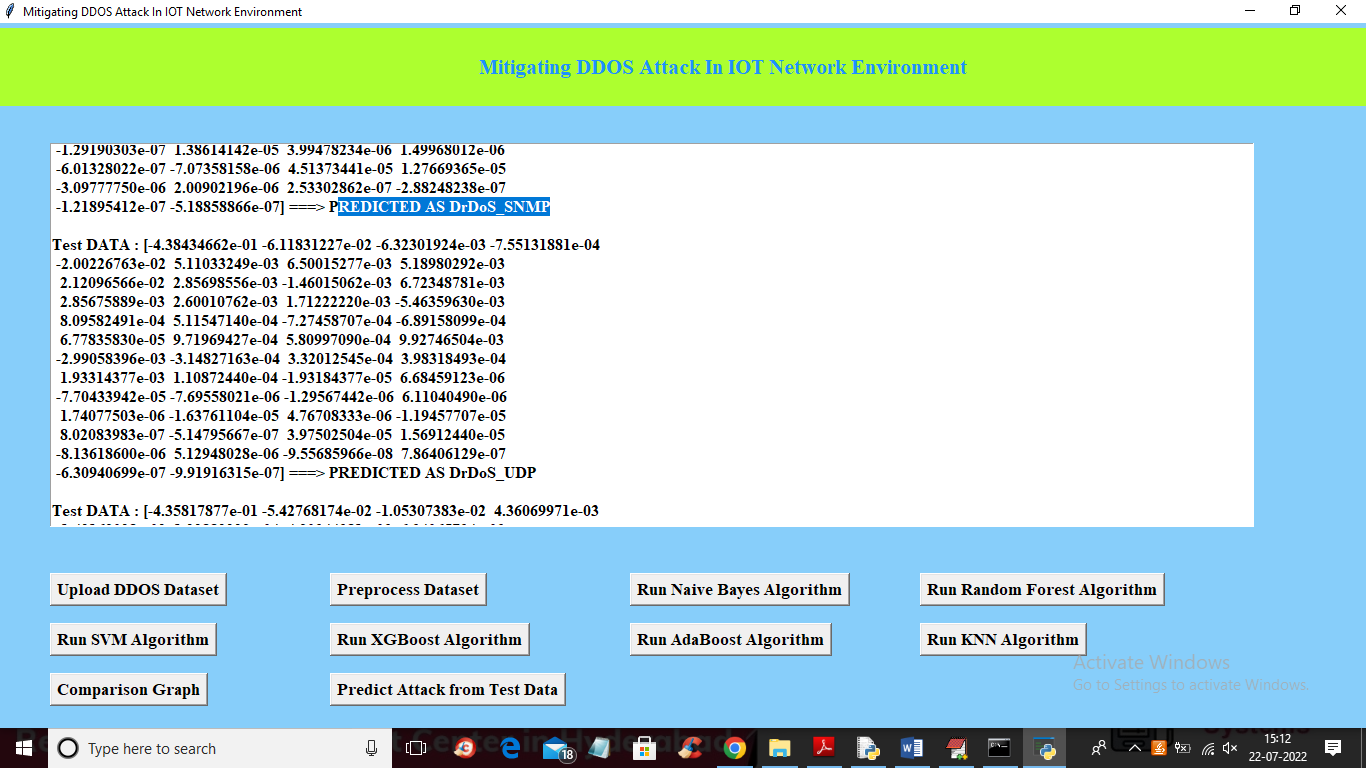
In above graph and comparison table we can see Random Forest got high accuracy and in above graph different colour bar represents different metrics such as accuracy, precision, recall and FSCORE. Now click on ‘Predict Attack from Test Data’ button to upload test data and get below output

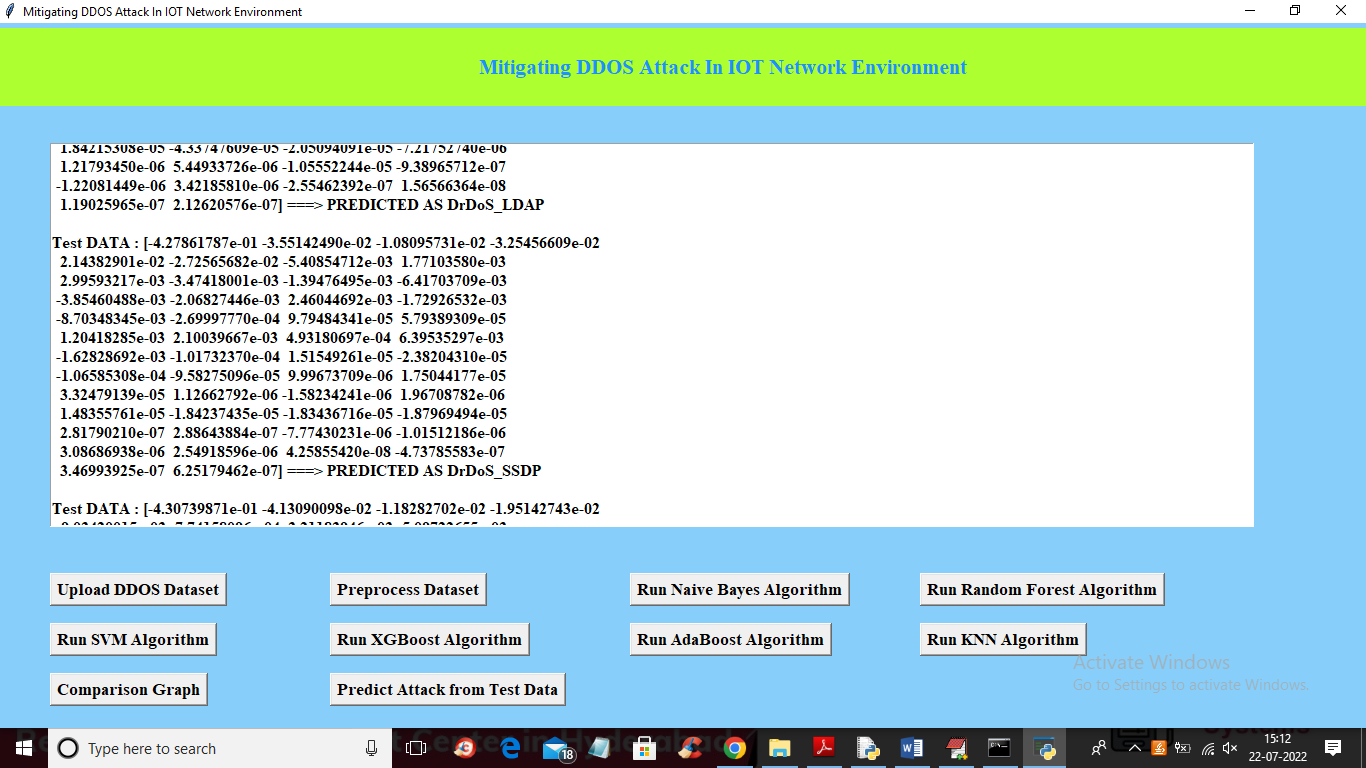


In above screen selecting and uploading TEST DATA file and then click on ‘Open’ button to get below output

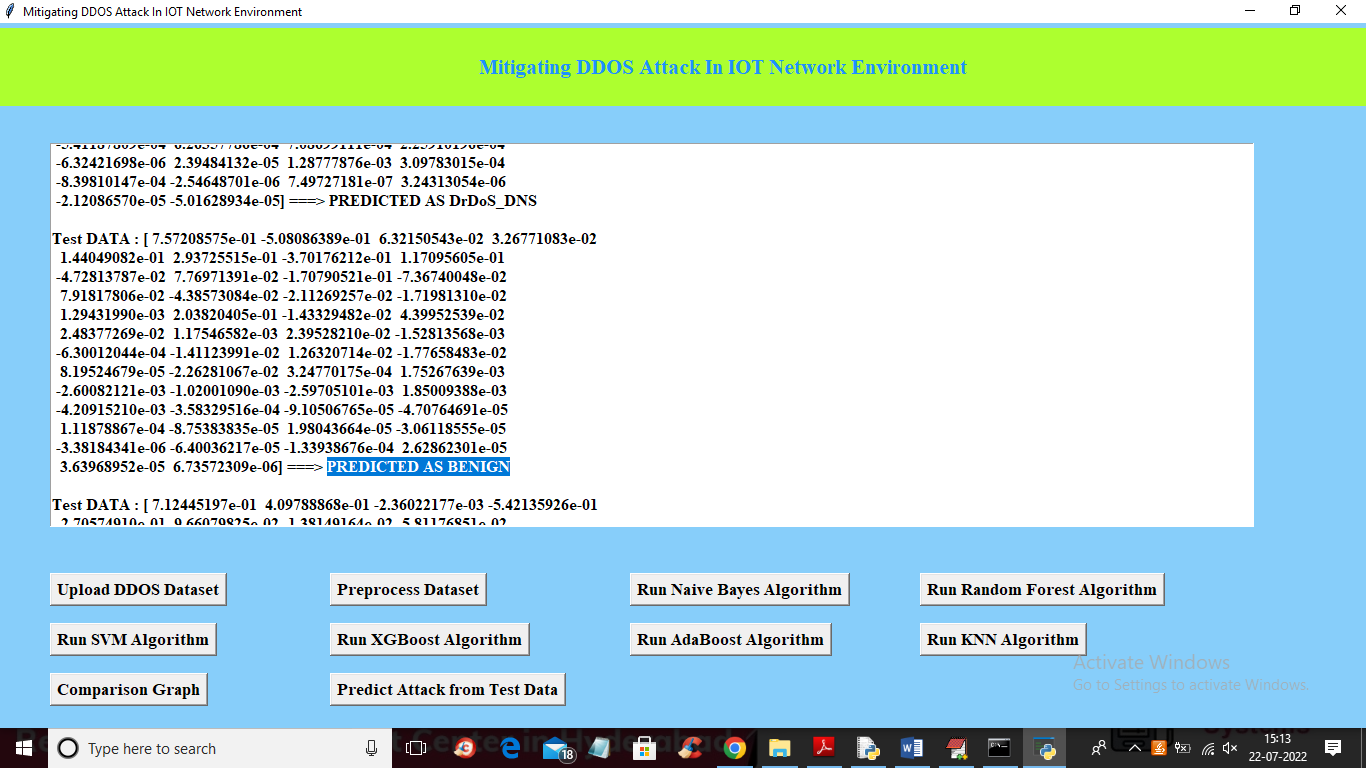


In above screen in square bracket we can see TEST DATA features and after arrow symbol =🡺 we can see predicted ATTACK as ‘SYN’ and scroll down above screen to view different predicted output









In above screens with each different test records different attacks and benign (normal) classes are predicted