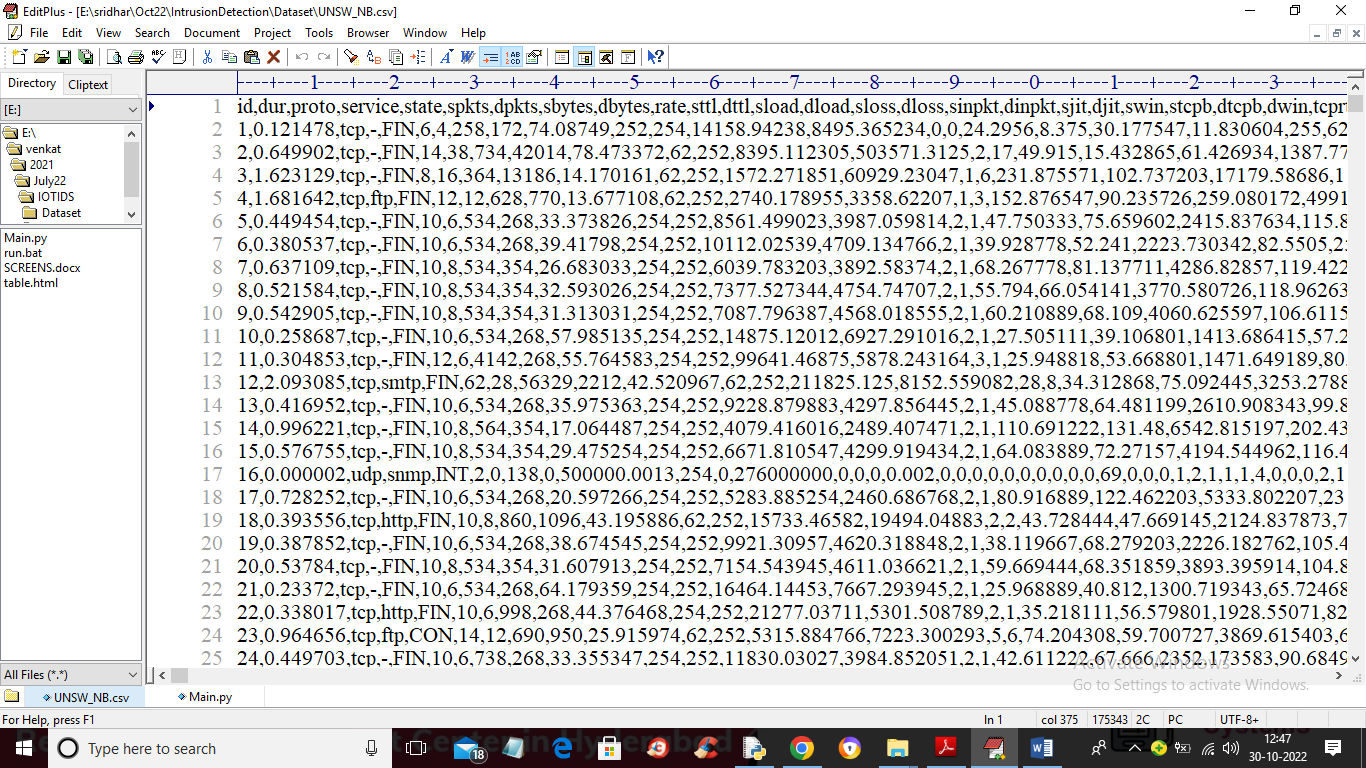
A Machine Learning-Based Classification and Prediction Technique for DDOS Attacks

In this paper author is employing Random Forest and XGBOOST algorithm to predict and classify DDOS attack using UNSW-15 dataset. Many existing algorithms are available to detect DDOS attack but they trained on old dataset called NSL-KDD and this dataset is not having information related to latest attack so they cannot predict attack with new signatures accurately.

To overcome from above problem author is training various machine learning algorithms called SVM, CNN, XGBOOST and Random Forest on latest UNSW-15 dataset and among all algorithms Random Forest and XGBOOST giving accuracy more than 90% and SVM performance is not up to the mark.

In propose paper author employing Preprocessing techniques to normalize dataset, label encoding, shuffling dataset, train and test split.

Propose UNSW dataset has information on nearly 10 different attacks and below screen showing dataset details



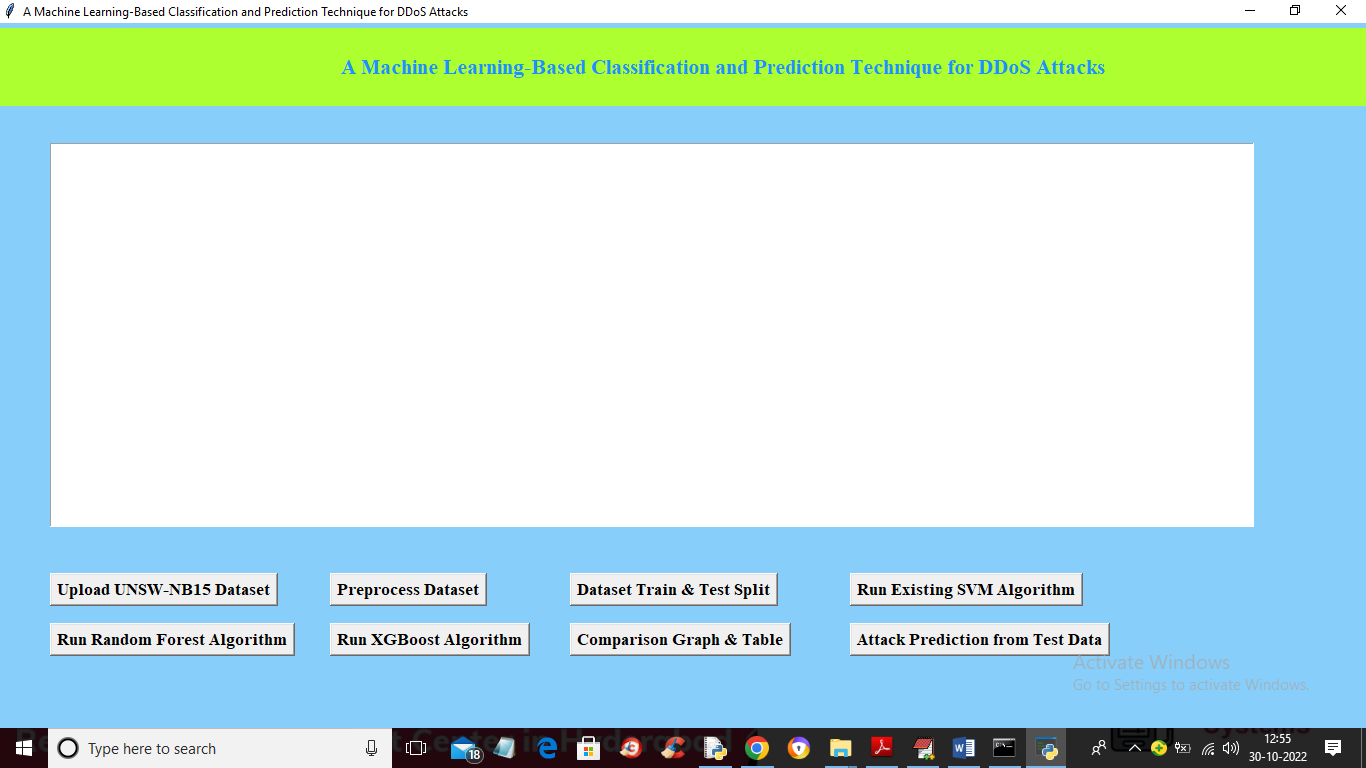
In above dataset screen first row contains dataset column names and remaining rows contains dataset values and all algorithms get trained on above dataset and their performance is evaluating in terms of accuracy, precision, recall, confusion matrix and FSCORE.

To implement this project we have designed following modules

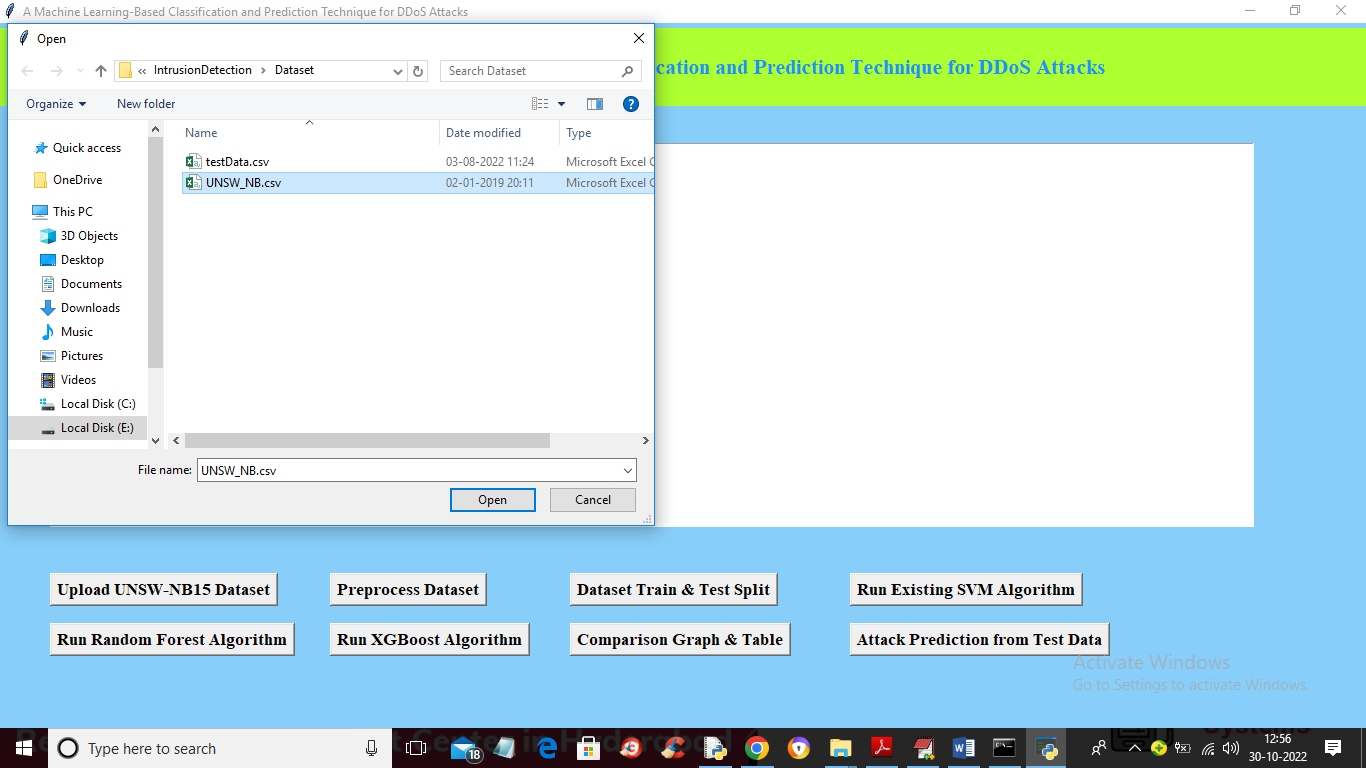
1. Upload UNSW-NB15 Dataset: using this module we will upload dataset to application and then will find and plot various attacks exists in dataset
2. Preprocess Dataset: using this module we will replace missing values and then convert non-numeric data to numeric data using label encoding algorithm. Normalize the dataset
3. Dataset Train & Test Split: using this module we will split dataset into train and test where application used 80% dataset for training and 20% for testing
4. Run Existing SVM Algorithm: using this module we will feed training data to SVM to train a model and this model will be applied on 20% test data to calculate prediction accuracy.
5. Run Random Forest Algorithm: using this module we will feed training data to Random Forest to train a model and this model will be applied on 20% test data to calculate prediction accuracy.
6. Run XGBoost Algorithm: using this module we will feed training data to XGBoost to train a model and this model will be applied on 20% test data to calculate prediction accuracy.
7. Comparison Graph & Table: using this module we will plot comparison graph between all algorithms
8. Attack Prediction from Test Data: using this module we will upload test data and then XGBOOST will analyse and perform prediction on test data

SCREEN SHOTS

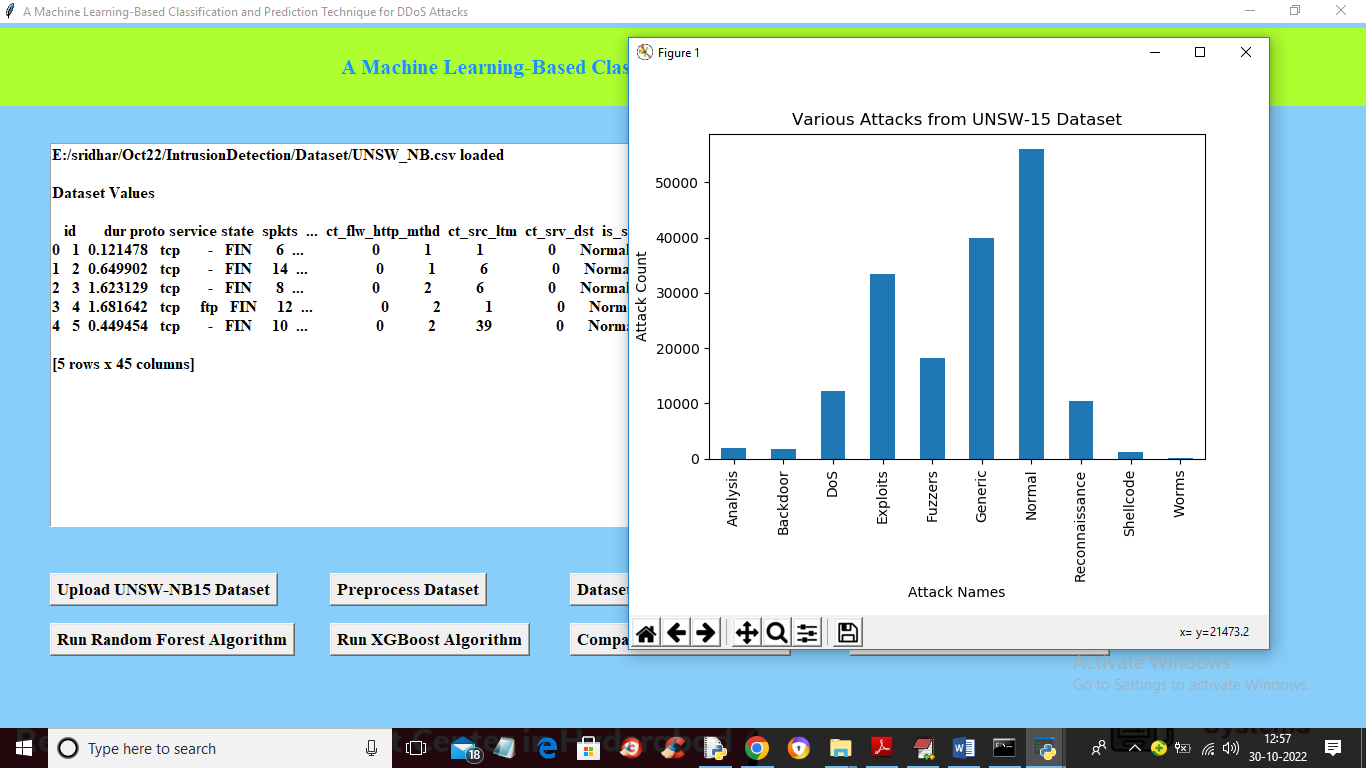
To run project double click on ‘run.bat’ file to get below screen



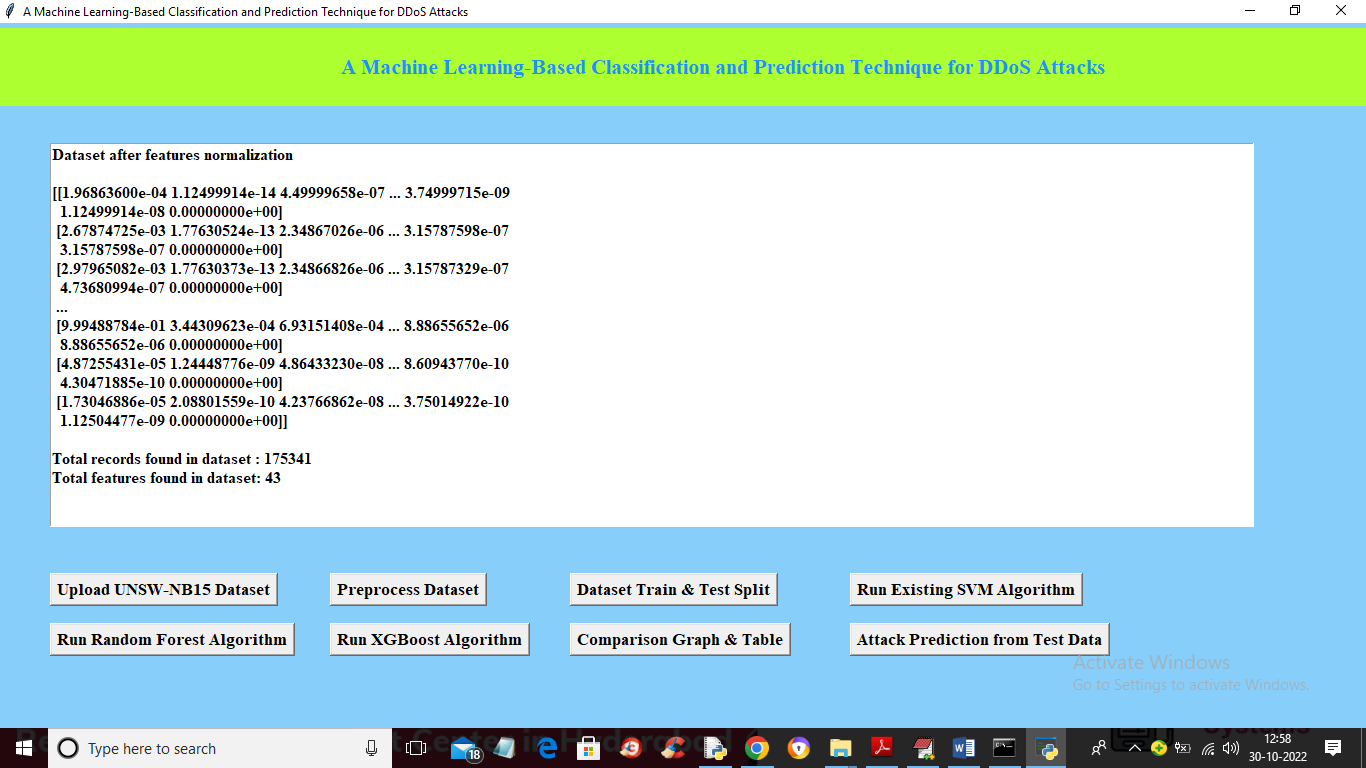
In above screen click on ‘Upload UNSW-15 Dataset’ button to upload dataset



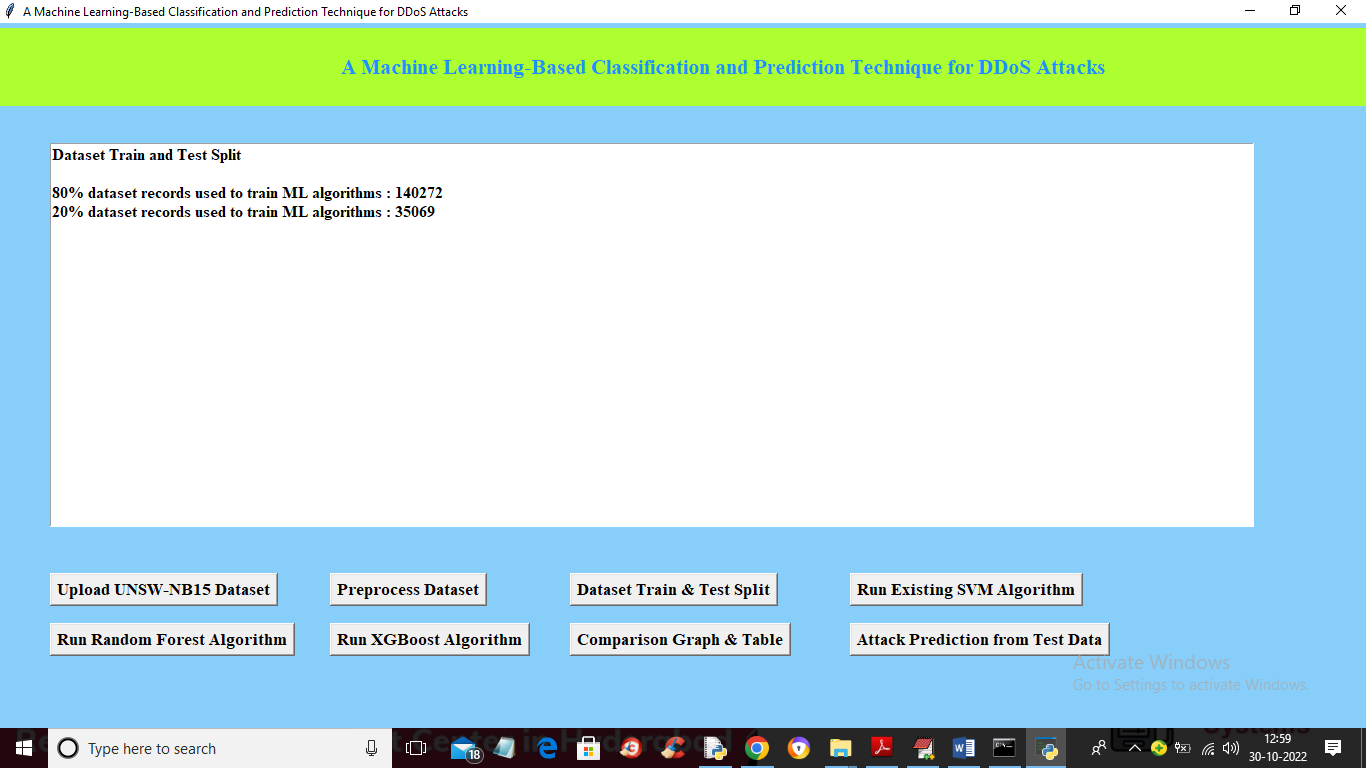
In above screen selecting and uploading ‘UNSW’ dataset and then click on ‘Open’ button to load dataset and get below output



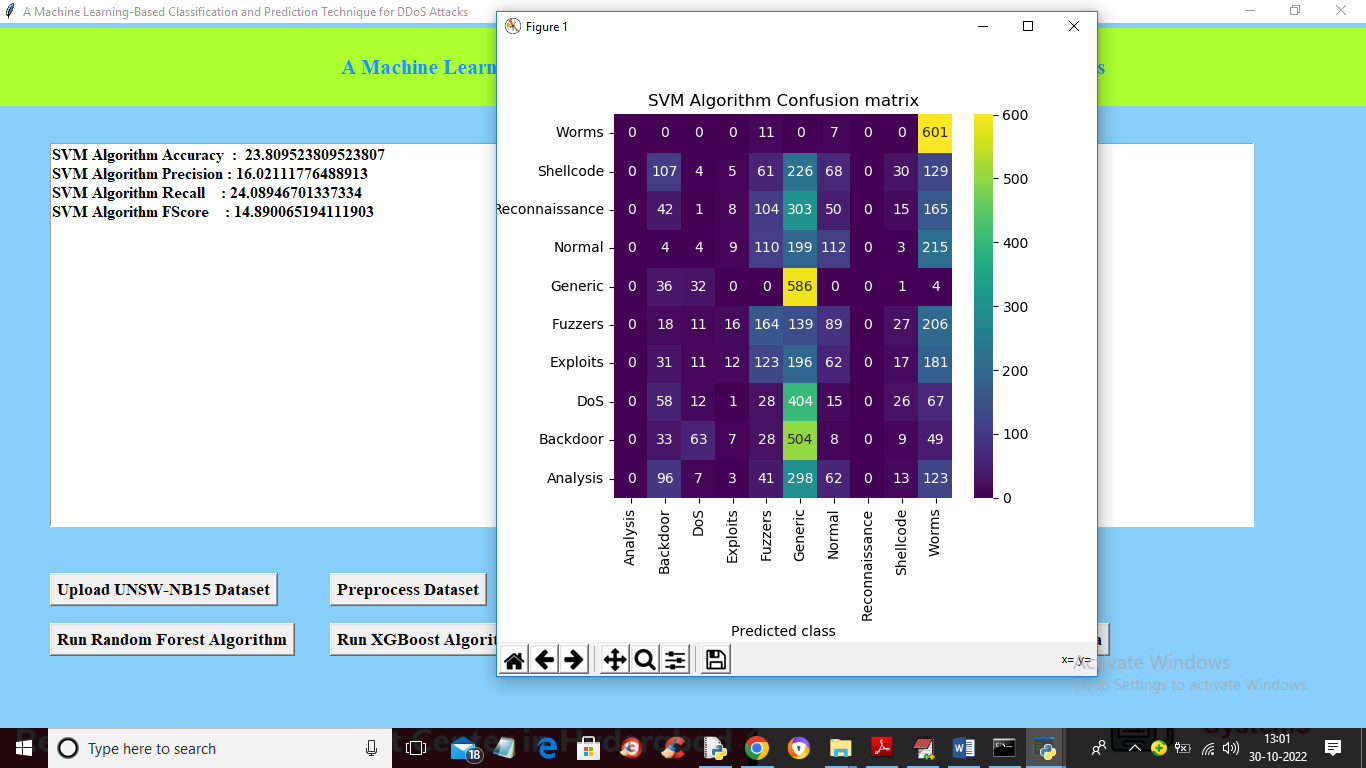
In above screen dataset loaded and we can see some values in the dataset are in non-numeric format and we have to process to convert to numeric and in graph x-axis contains ATTACK NAMES and y-axis contains its count in dataset and now close above graph and then click on ‘Preprocess Dataset’ button to get below output



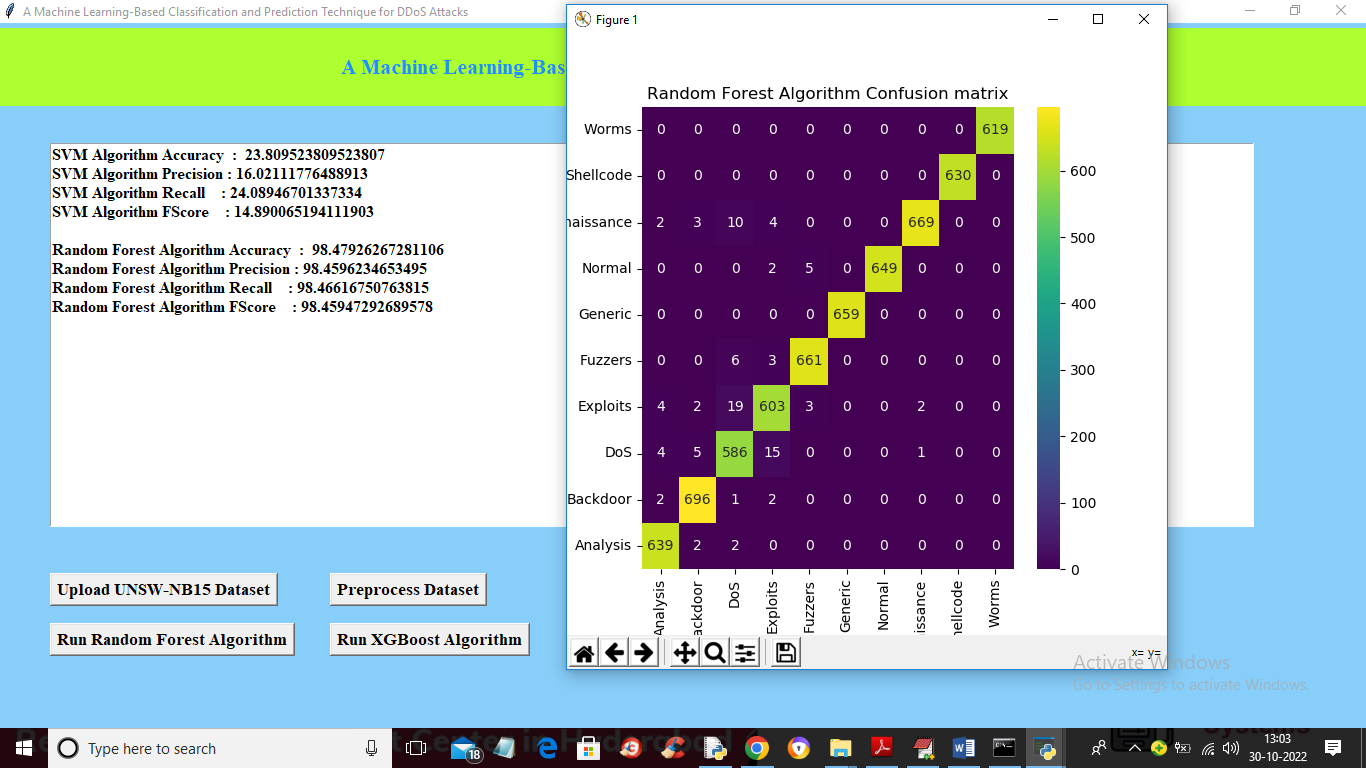
In above screen we can see entire dataset converted to numeric format and then in last 2 lines we can see dataset and its column size and now click on ‘Dataset Train & Test Split’ button to split dataset and get below output



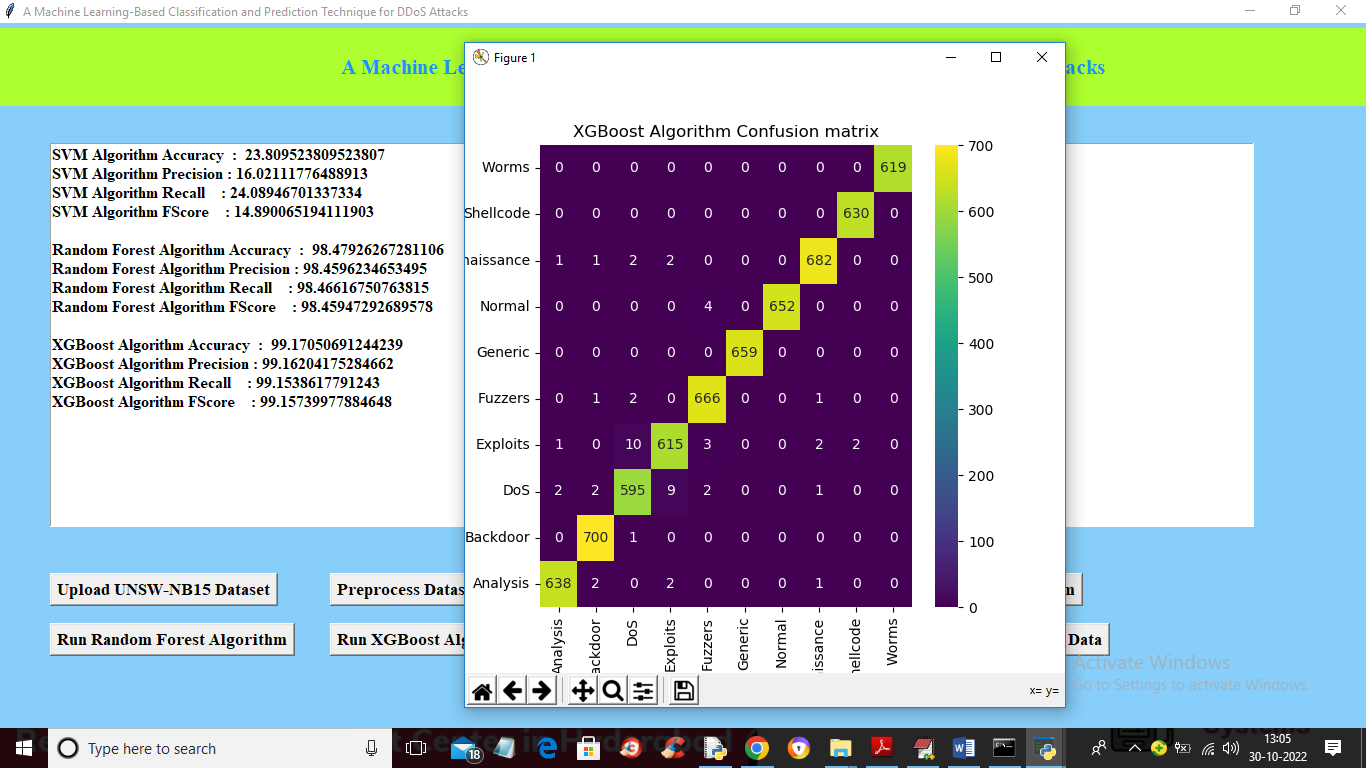
In above screen we can see dataset size used for training and test and now click on ‘Run Existing SVM Algorithm’ button to train SVM and get below output



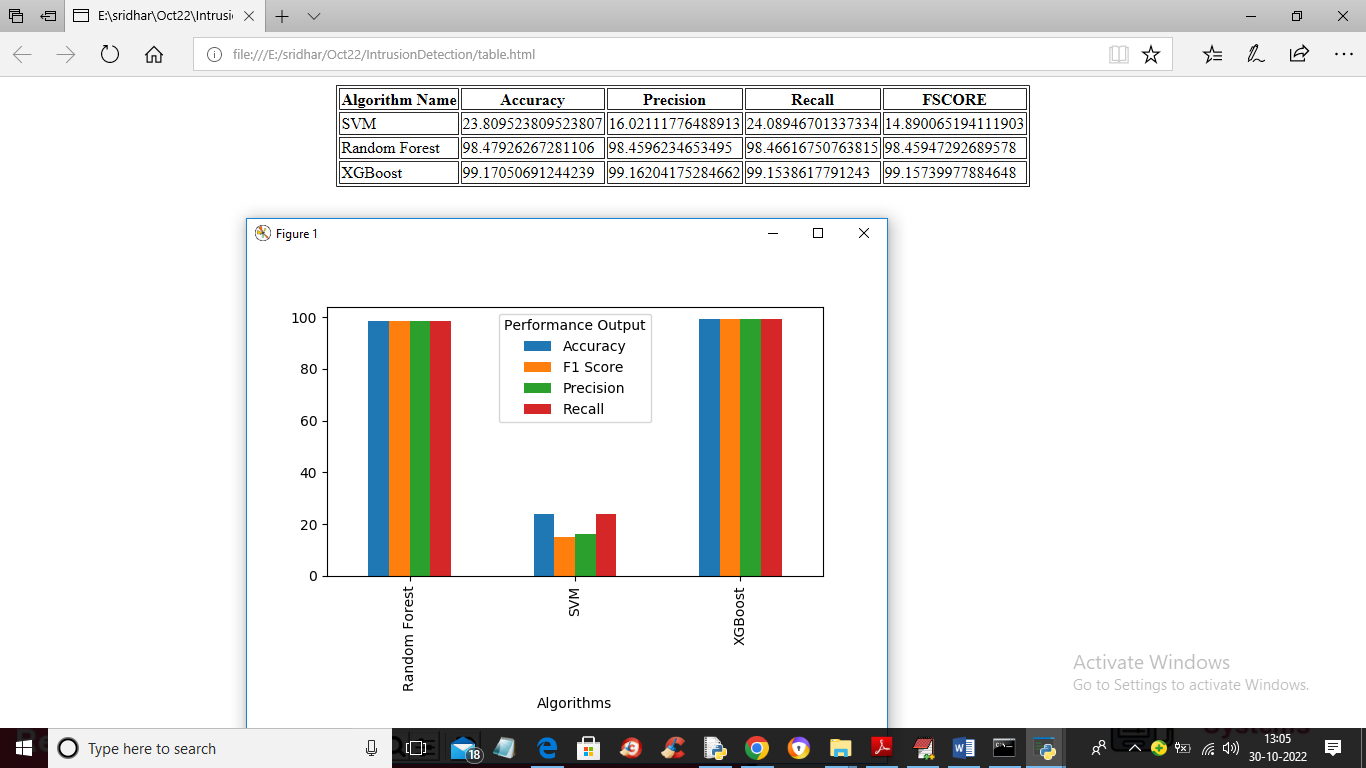
In above screen with SVM we got 23% accuracy and in confusion matrix graph x-axis represents Predicted classes and y-axis represents TRUE colour classes and in above graph all counts in blue colour boxes are wrong prediction and different colour boxes represents correct prediction count and we can see SVM predicted most of the records as incorrect and now close above graph and then click on ‘Run Random Forest Algorithm’ button to train random forest and get below output



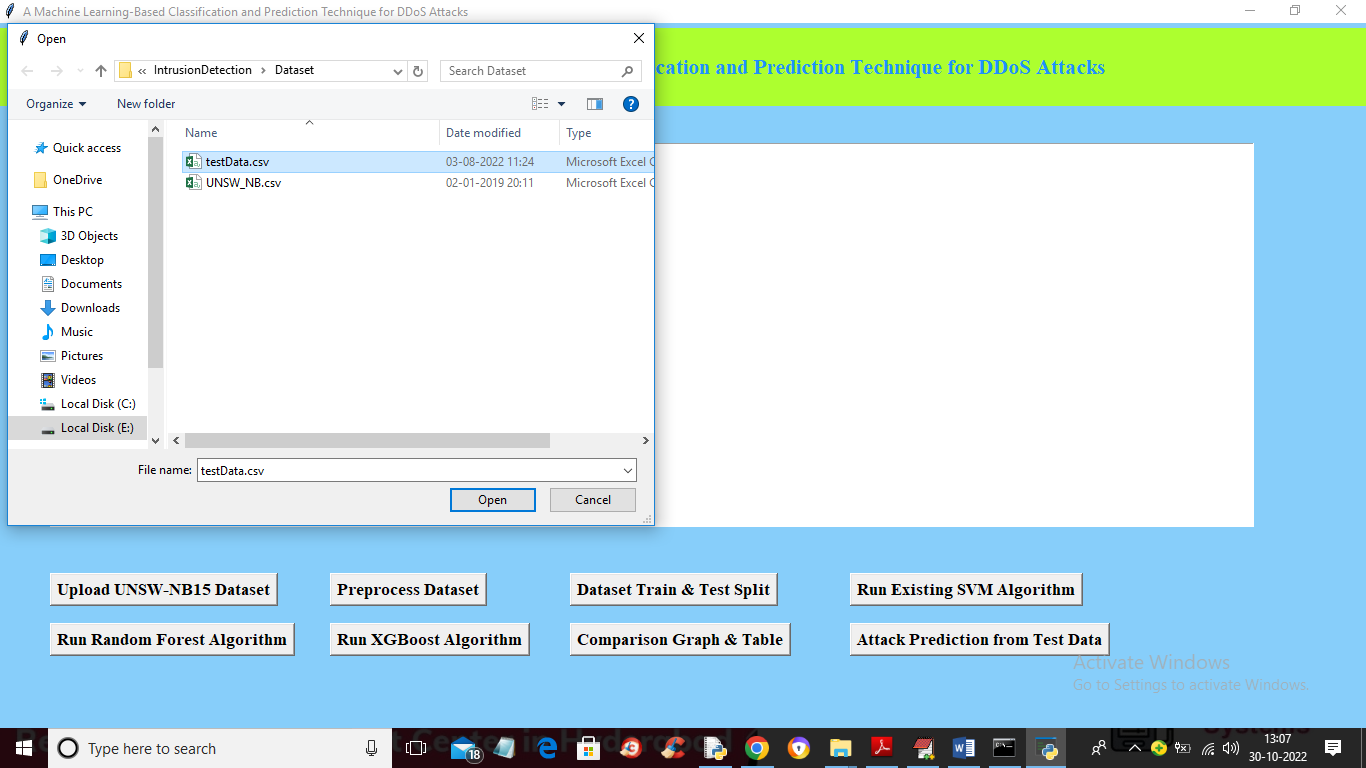
In above screen with Random Forest we got 98% accuracy and blue colour boxes in confusion matrix are wrong prediction and we can see colour boxes contains correct prediction count and random forest predicted more than 95% records correctly. Now click on ‘Train XGBOOST’ button to get below output



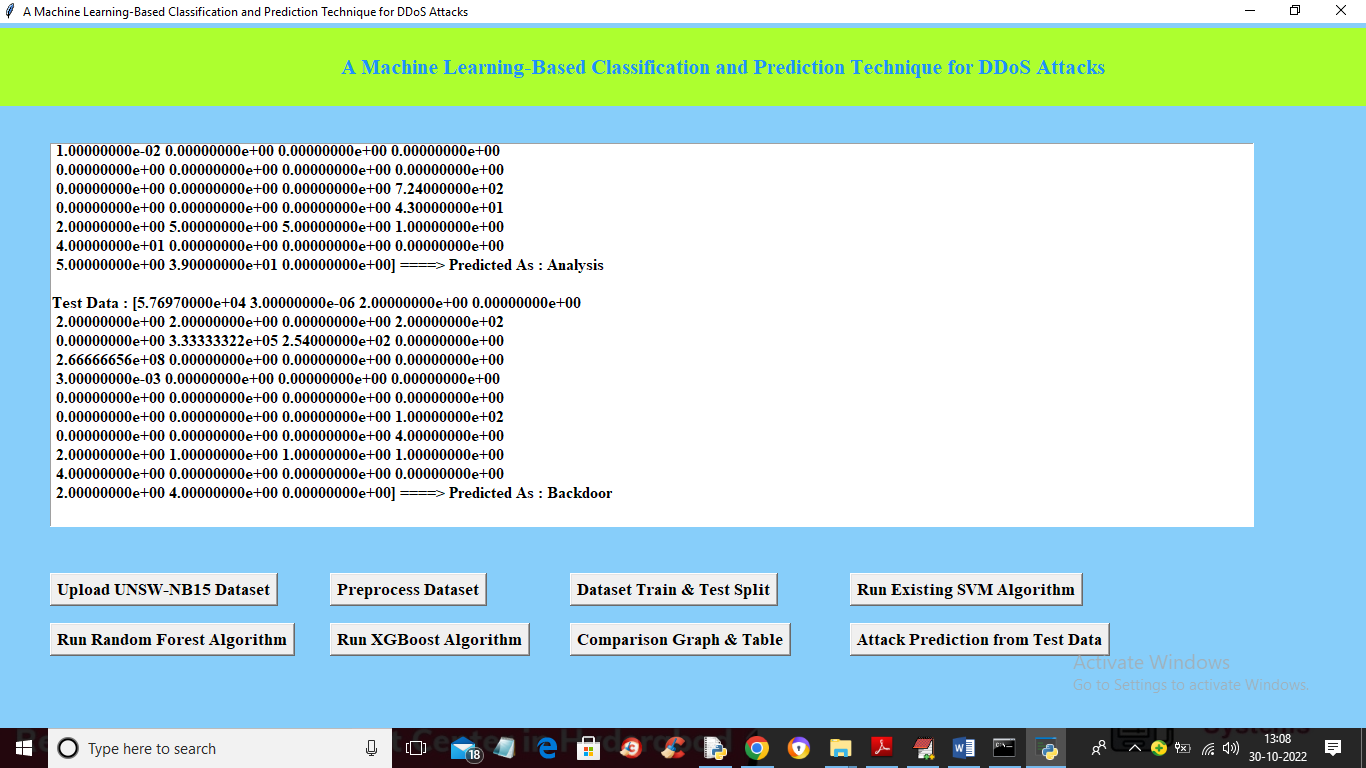
In above screen with XGBOOST we got 99% accuracy and now click on ‘Comparison Graph’ button to get below graph



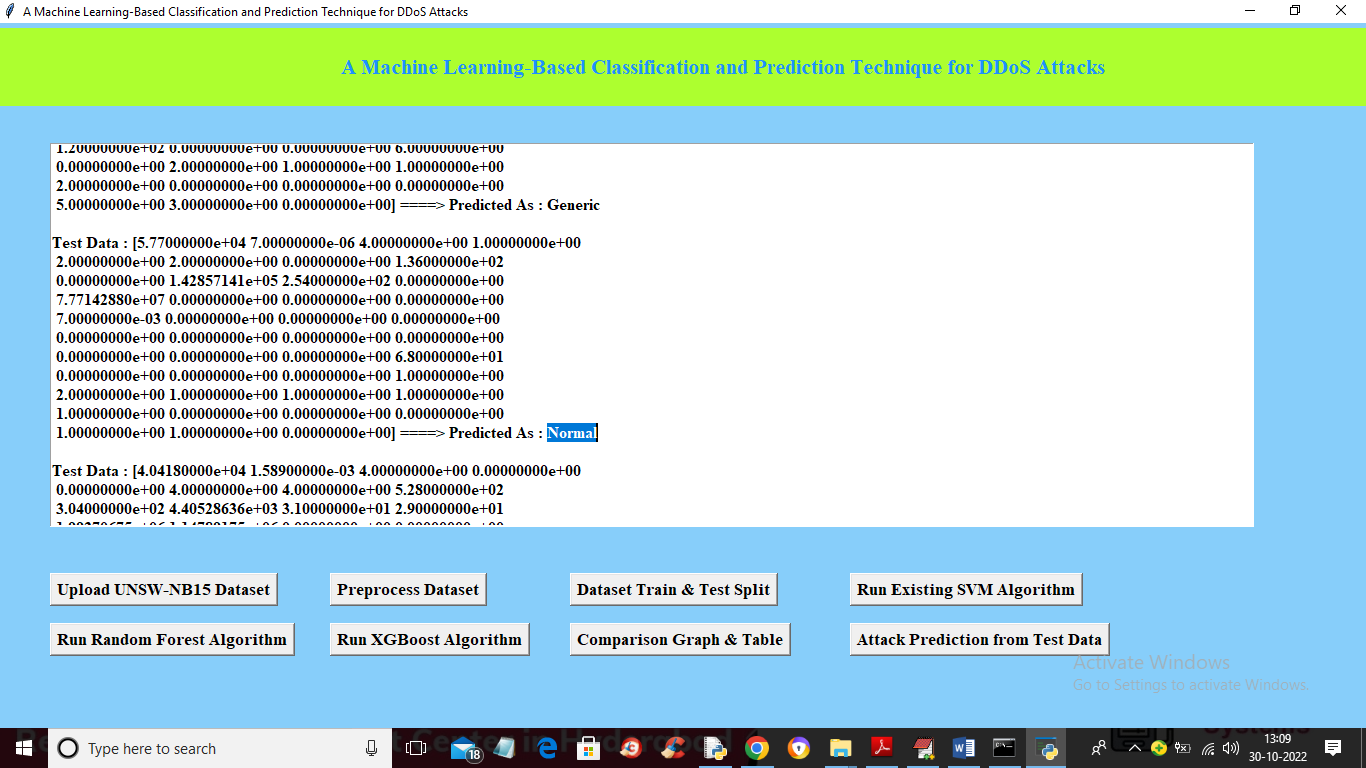
In above graph x-axis represents algorithm names with different colour bar represents different metrics such as Accuracy, precision etc. Y-axis represents values and same performance of algorithms we can see in tabular format and in all algorithms XGBOOST got best accuracy and now click on ‘Attack Prediction from Test Data’ button to upload test data and get below output

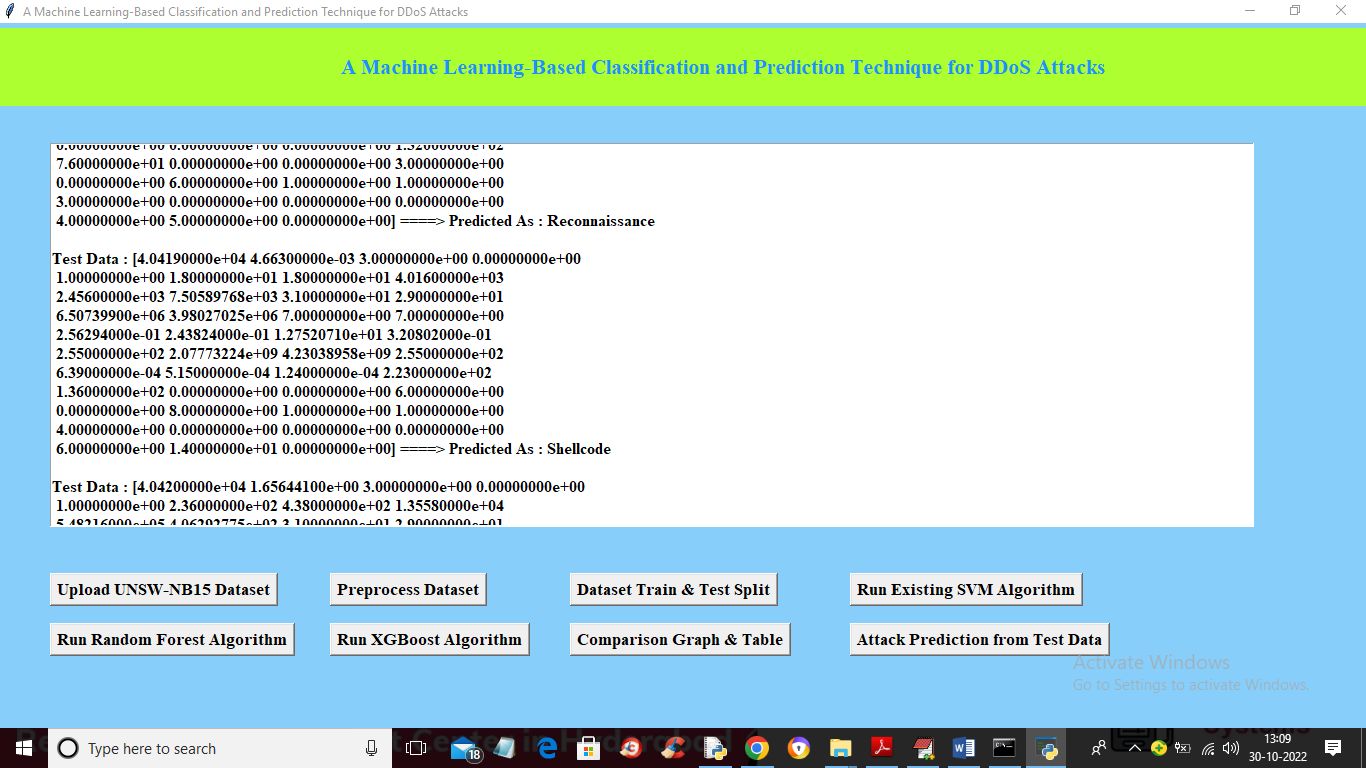


In above screen selecting and uploading ‘testData.csv’ file and then click on ‘Open’ button to get below output



In above screen in square bracket we can see TEST data and after array symbol we can see attack prediction name and you can scroll down above screen to view all predicted attacks





So you can see different predicted attacks