Network Traffic Analysis Using Machine Learning

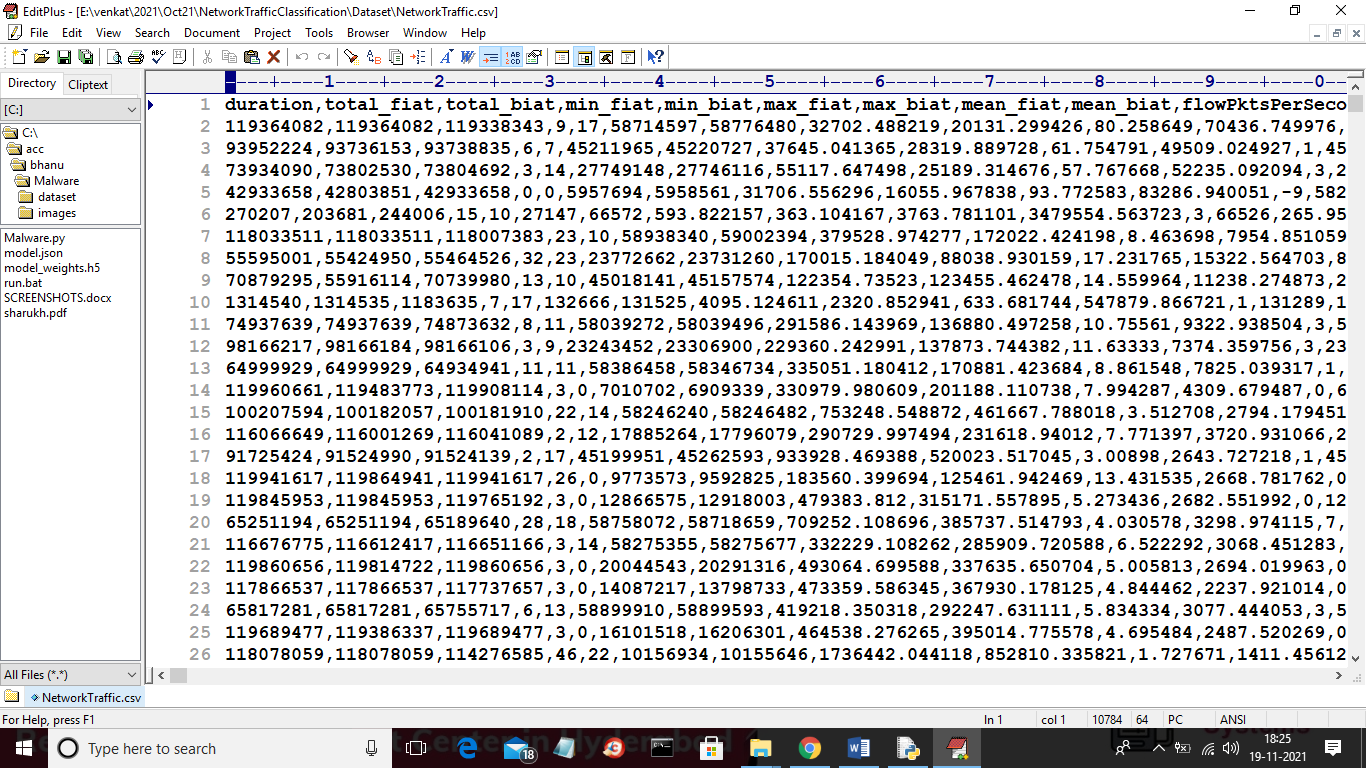
In this project we are using various machine learning algorithms such as KNN, SVM, Decision Tree and Naïve Bayes to predict traffic or classify type of network data such as BROWSING traffic, MAIL traffic etc. Lots of network traffic type of data is available but in this project we are training ML algorithms to predict or classify 14 different types of traffic. Below are the names of such traffic types

['BROWSING', 'CHAT', 'FT', 'MAIL', 'P2P', 'STREAMING', 'VOIP', 'VPN-BROWSING', 'VPN-CHAT', 'VPN-FT', 'VPN-MAIL', 'VPN-P2P', 'VPN-STREAMING', 'VPN-VOIP']

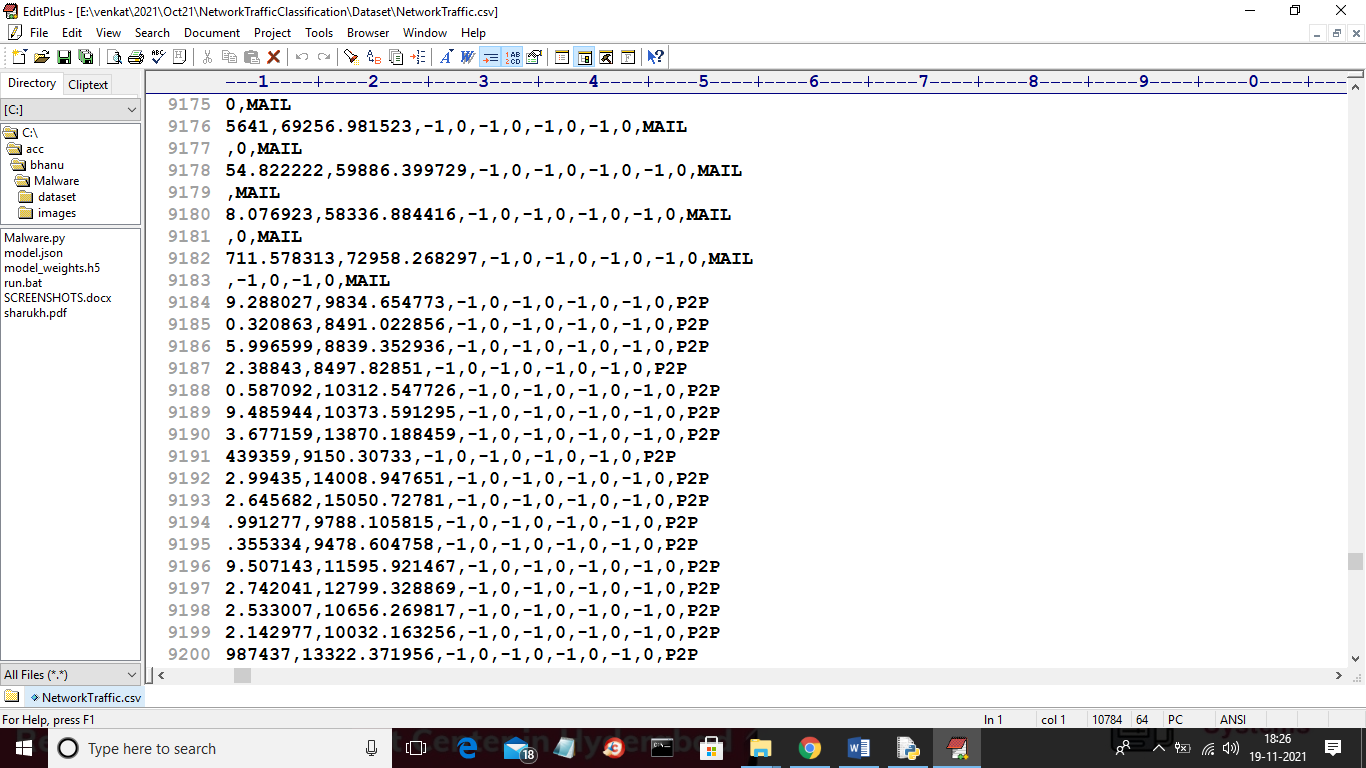
We have evaluated performance of each algorithm in terms of accuracy, precision, recall and FSCORE and in all algorithms KNN shows better result. We downloaded dataset from below link

<https://www.unb.ca/cic/datasets/vpn.html>

below is the dataset screen shot



In above dataset first row represents dataset column names and remaining rows represents dataset values and in last column we can see type of traffic which you can see in below screen



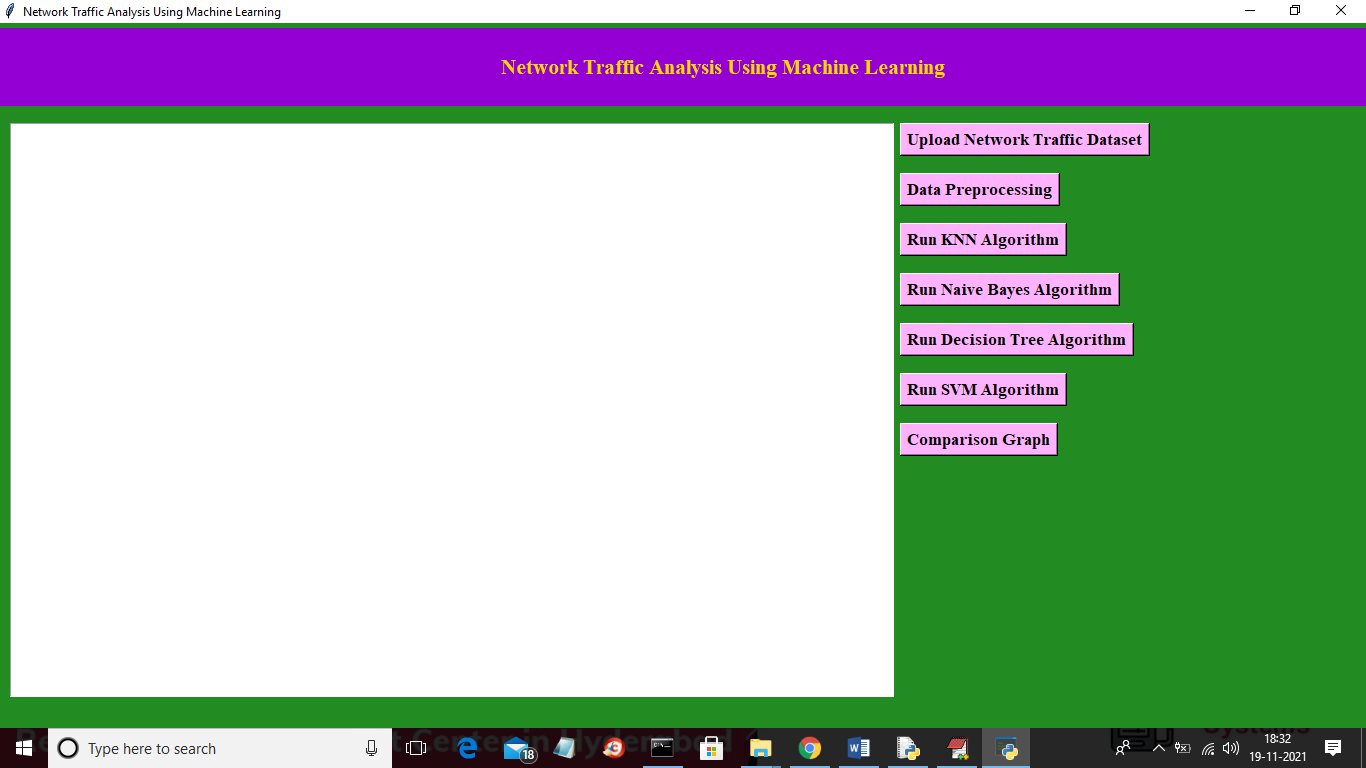
In last column you can see each record is associated with one label such as traffic type.

To implement this project we have designed following modules

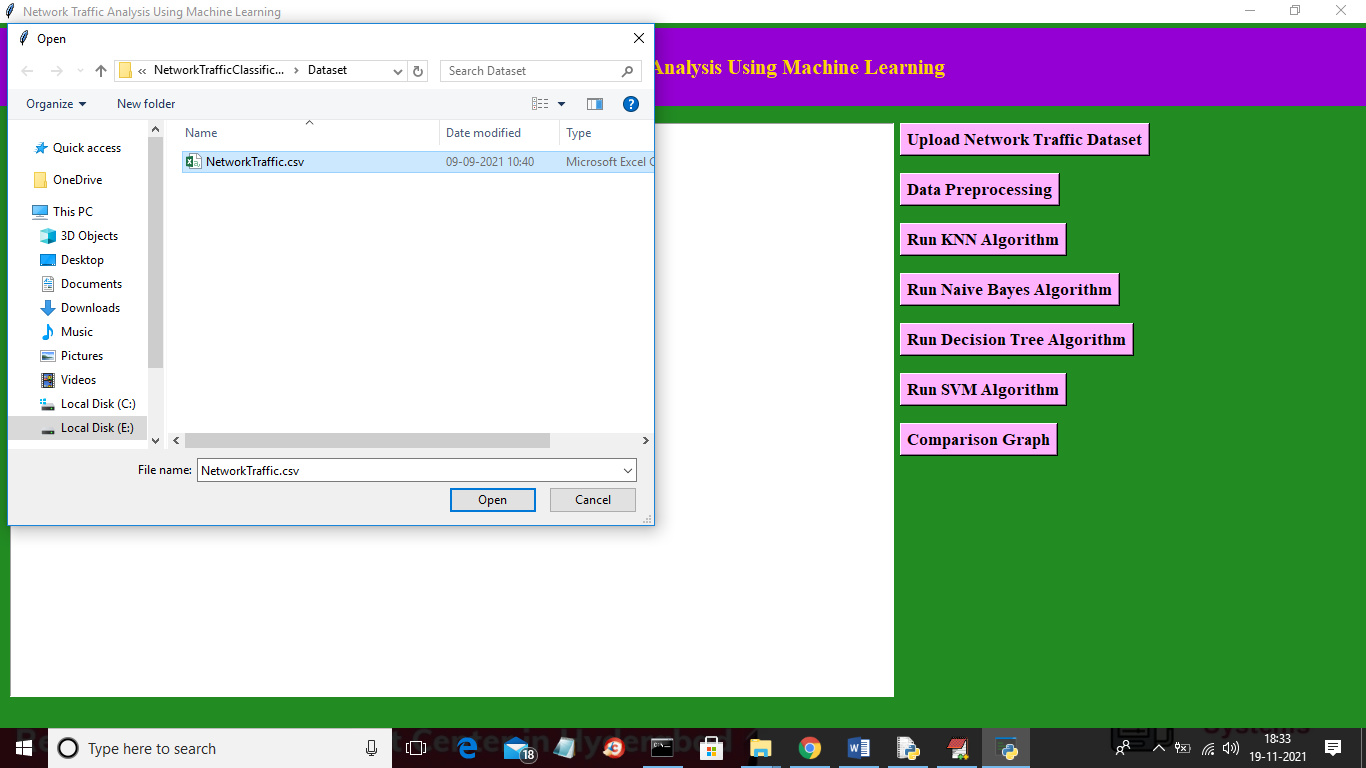
1. Upload Network Traffic Dataset: using this module we will upload dataset to application
2. Data Preprocessing: using this module we will clean dataset by removing missing values and ML algorithms will not take characters values so we need to process dataset to convert all non-numeric characters to numeric characters by assigning integer ID to each unique string data. After processing we will split dataset into train and test where 80% dataset will be used for training and 20% for testing
3. Run KNN Algorithm: using above train data we will train KNN algorithm and this trained model will be applied on test data to predict traffic type and then calculate accuracy of correct prediction
4. Run Naive Bayes Algorithm: using above train data we will train Naïve Bayes algorithm and this trained model will be applied on test data to predict traffic type and then calculate accuracy of correct prediction
5. Run Decision Tree Algorithm: using above train data we will train Decision tree algorithm and this trained model will be applied on test data to predict traffic type and then calculate accuracy of correct prediction
6. Run SVM Algorithm: using above train data we will train SVM algorithm and this trained model will be applied on test data to predict traffic type and then calculate accuracy of correct prediction
7. Comparison Graph: using this module we will show accuracy graph between all algorithms

SCREEN SHOTS

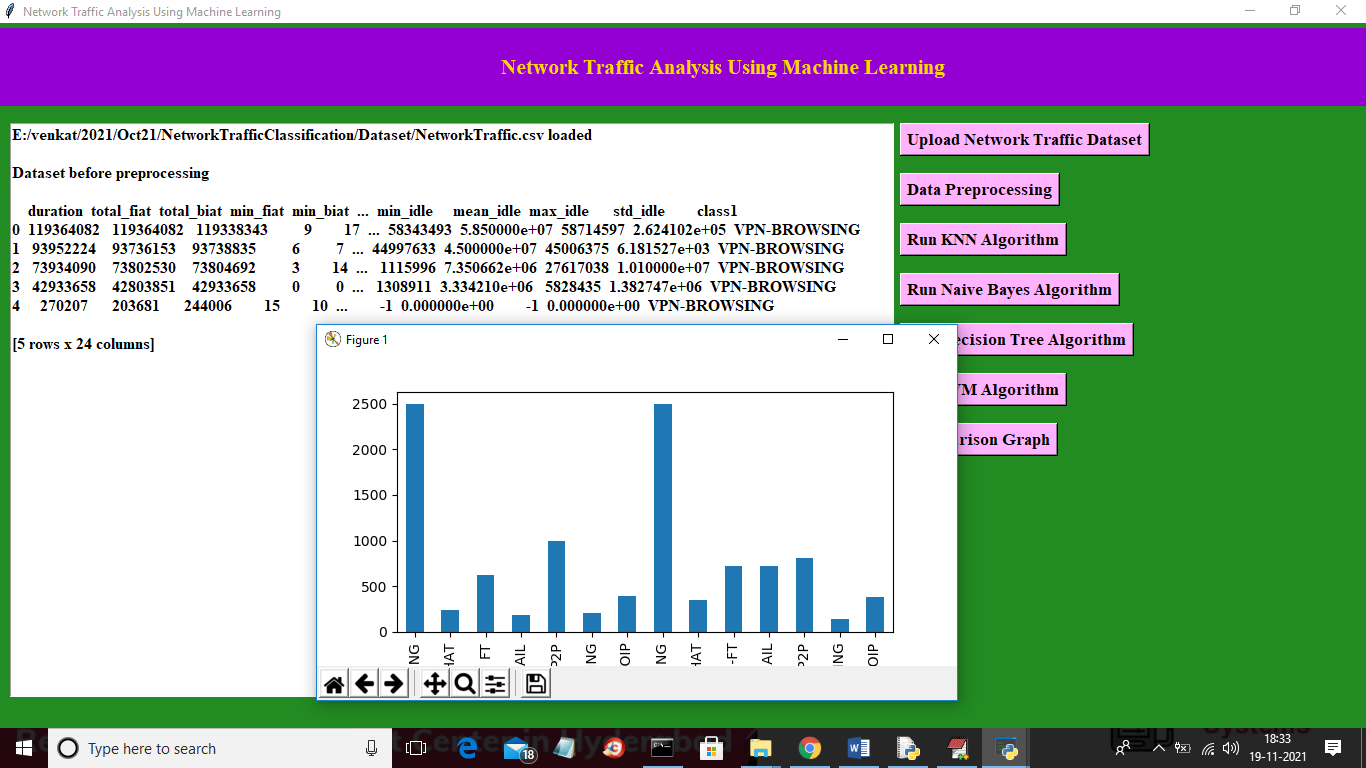
To run project double click on ‘run.bat’ file to get below screen



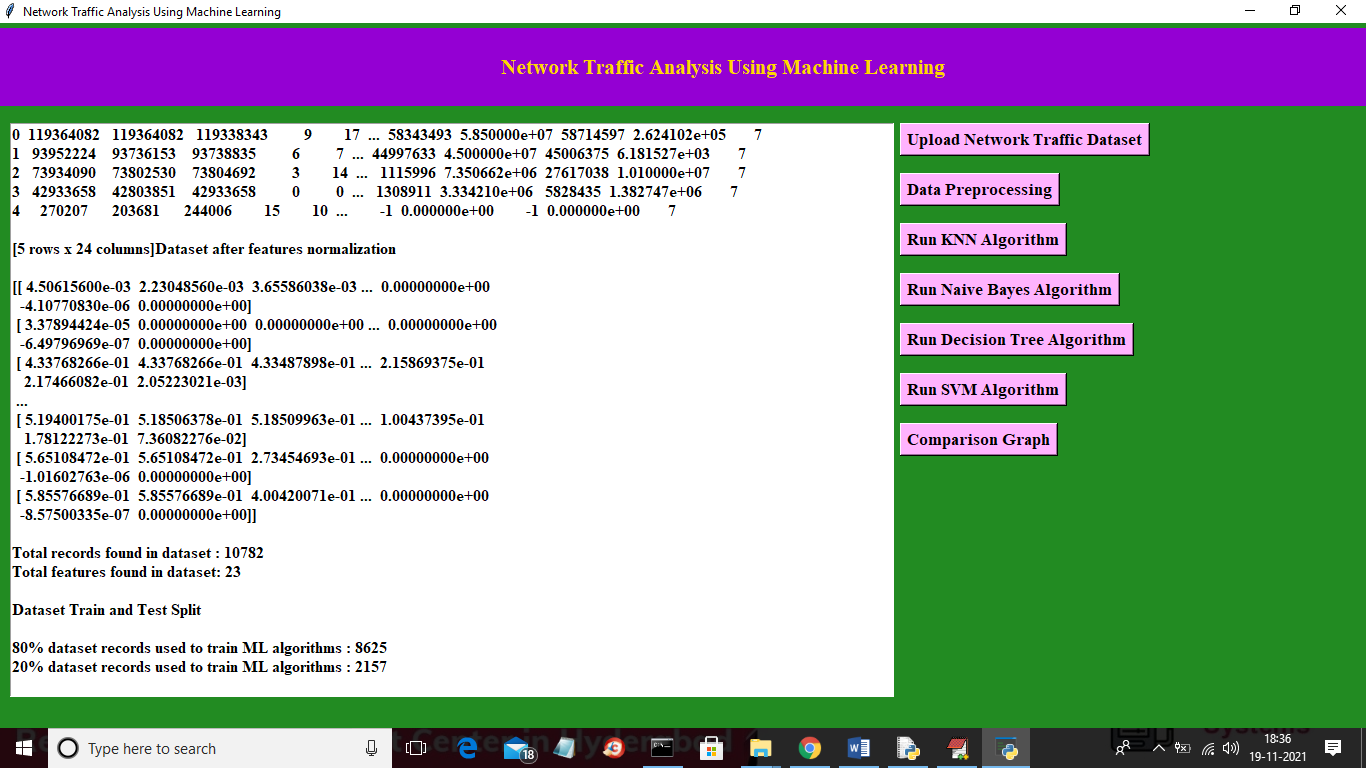
In above screen click on ‘upload Network Traffic Dataset’ button to upload dataset



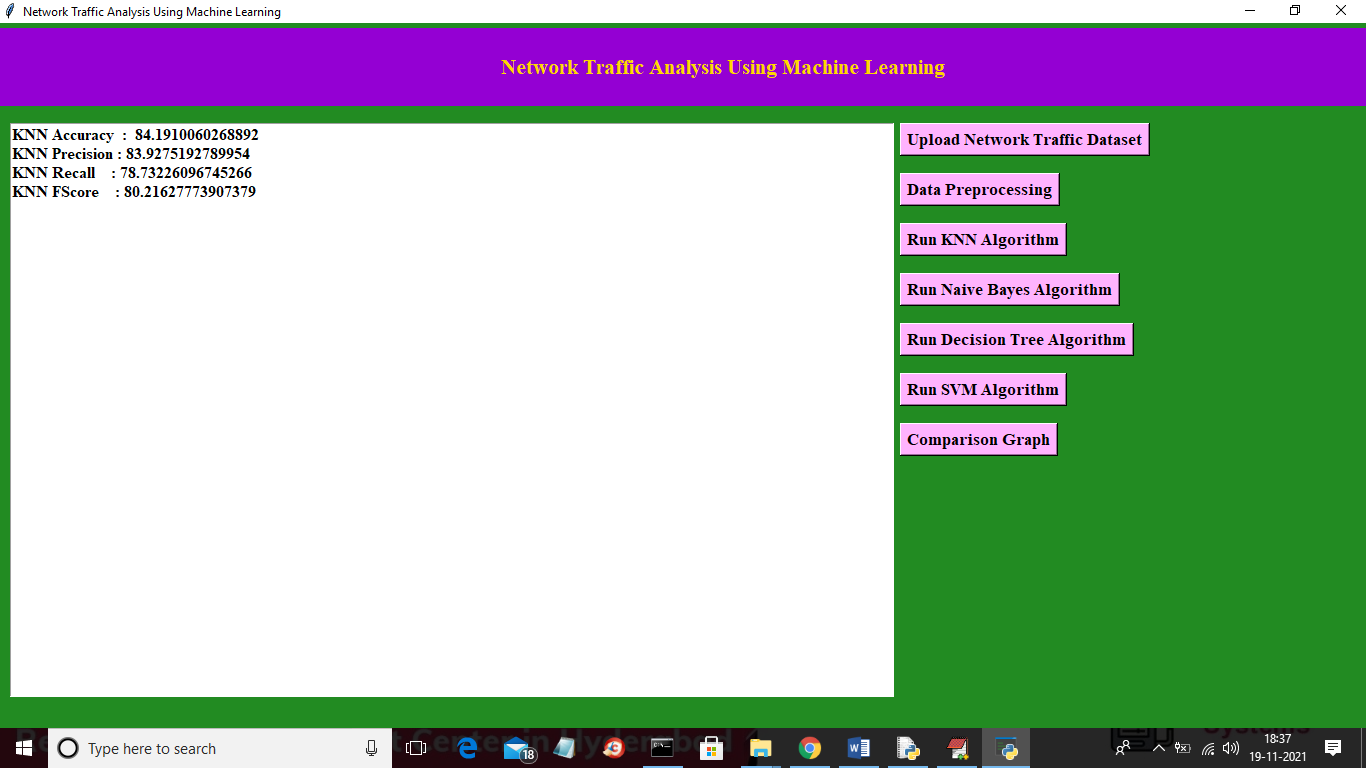
In above screen selecting and uploading ‘NetworkTraffic.csv’ file and then click on ‘Open’ button to load dataset and to get below screen



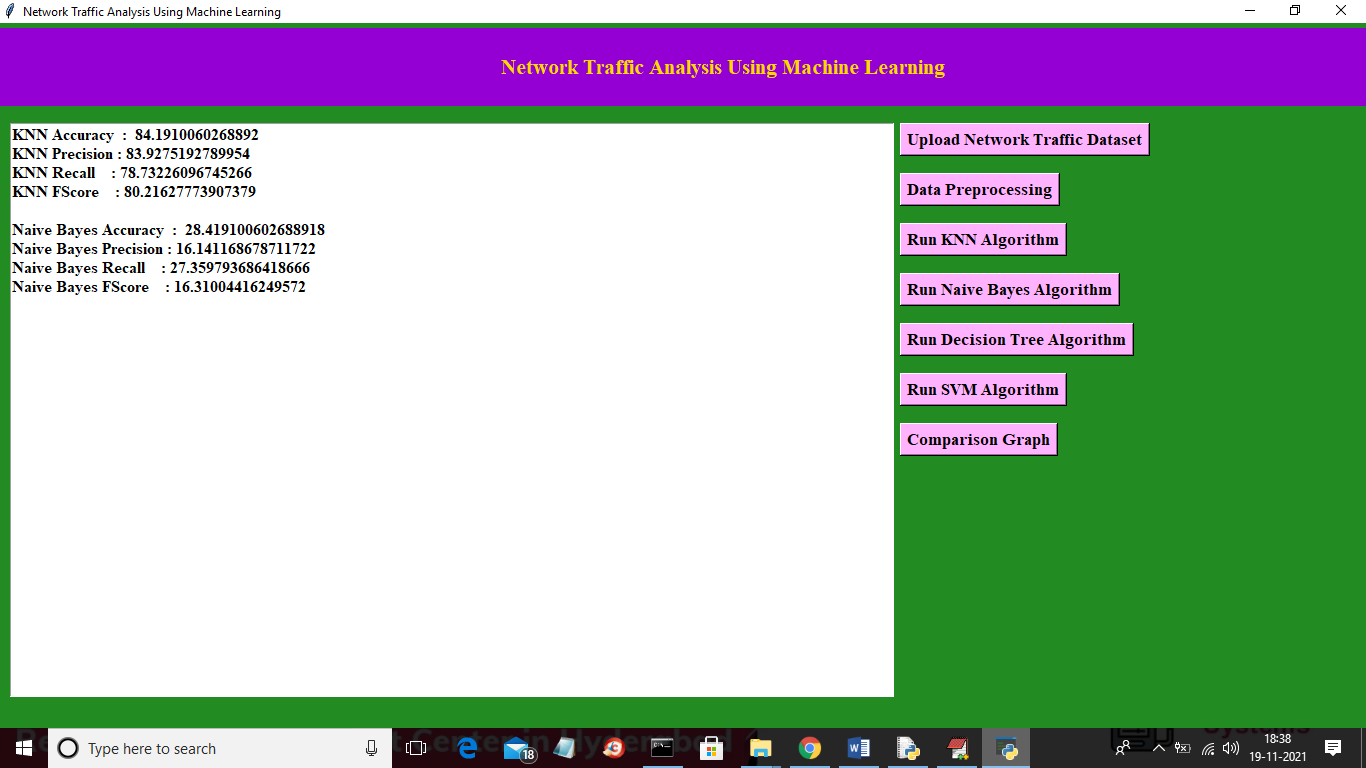
In above screen we can see dataset loaded and dataset contains lots of non-numeric values so we need to process it and in graph x-axis we can see traffic type and y-axis represents total records in dataset for that traffic. Now close above graph and then click on ‘Data Preprocessing’ to clean dataset



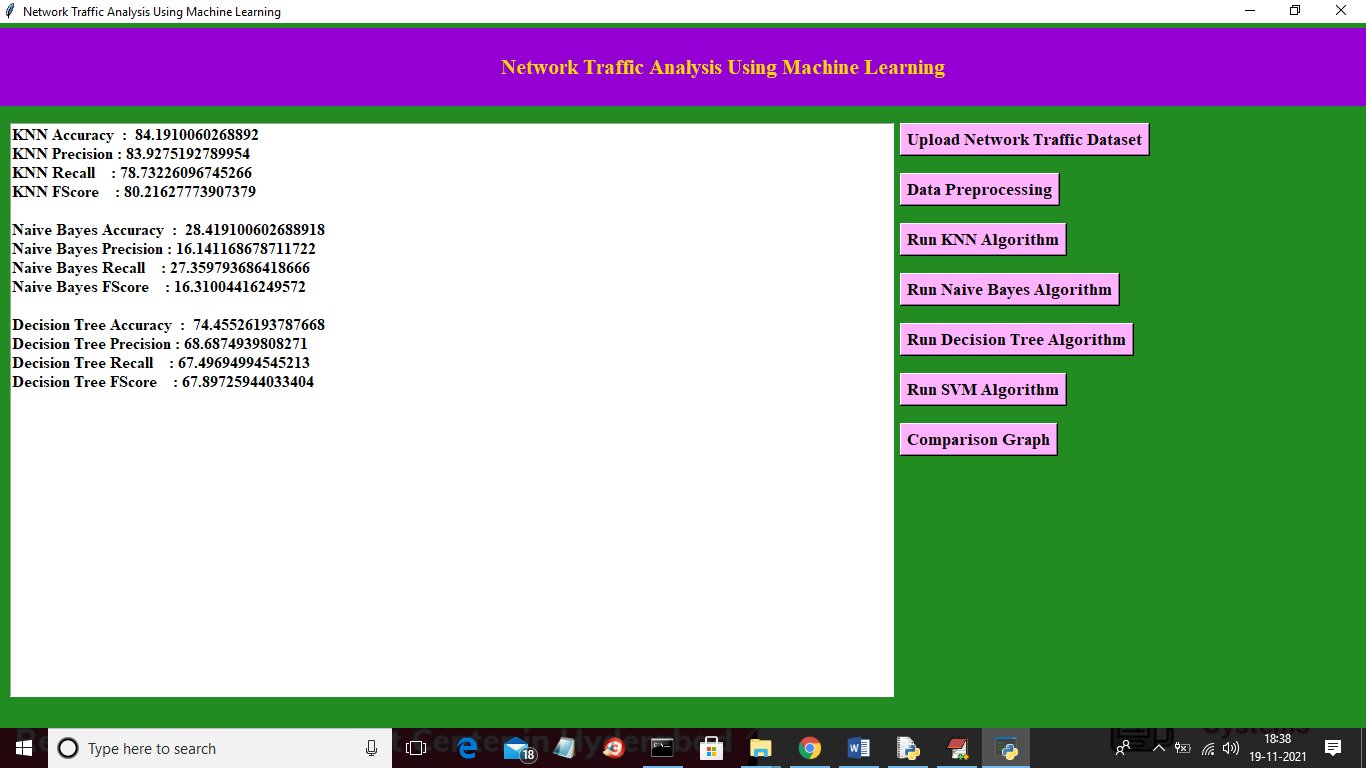
In above screen we can see all dataset converted to numeric format and in last we can see total records and columns found in dataset and then splitted dataset percentage for train and test records. Now train and test data is ready and now click on ‘Run KNN Algorithm’ button to train KNN and get below result



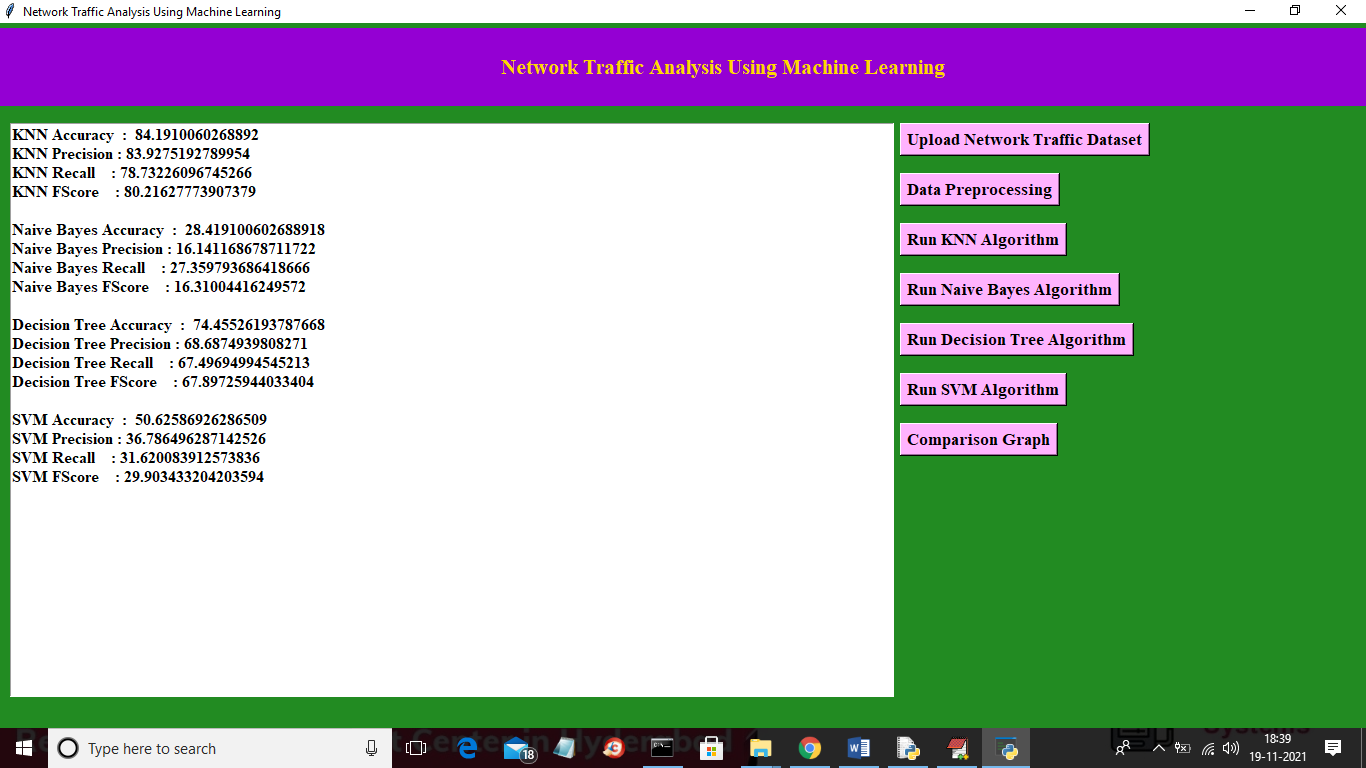
In above screen with KNN we got 84% accuracy and now click on ‘Run Naïve Bayes Algorithm’ button to train it



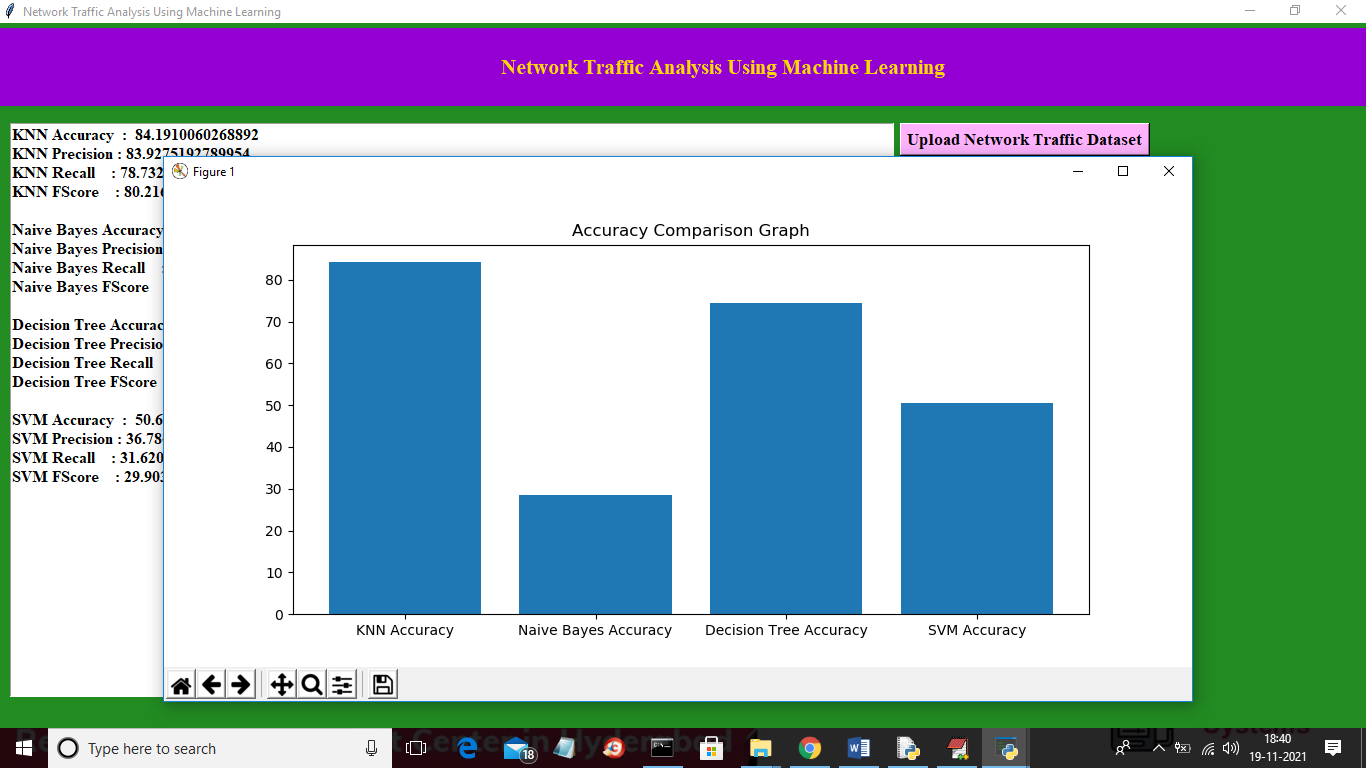
In above screen for same dataset with naïve bayes we got 28% accuracy and now click on ‘Run Decision Tree Algorithm’ button to get below result



In above screen for same dataset with SVM we got 74% accuracy and now click on ‘Run SVM Algorithm’ button to get below result



In above screen with SVM we got 50% accuracy and now click on ‘Comparison Graph’ button to get below result



In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and in all algorithms KNN shows better result