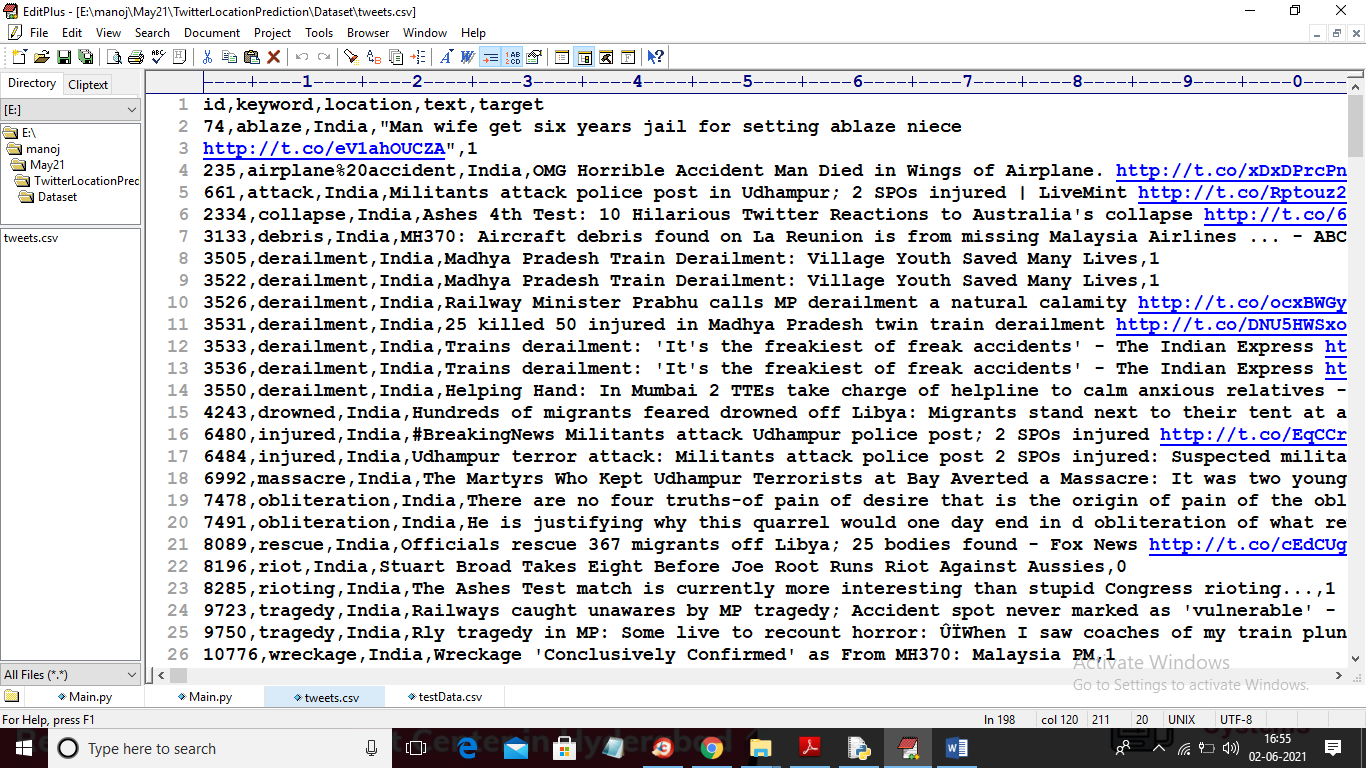
Location prediction on Twitter using machine learning Techniques

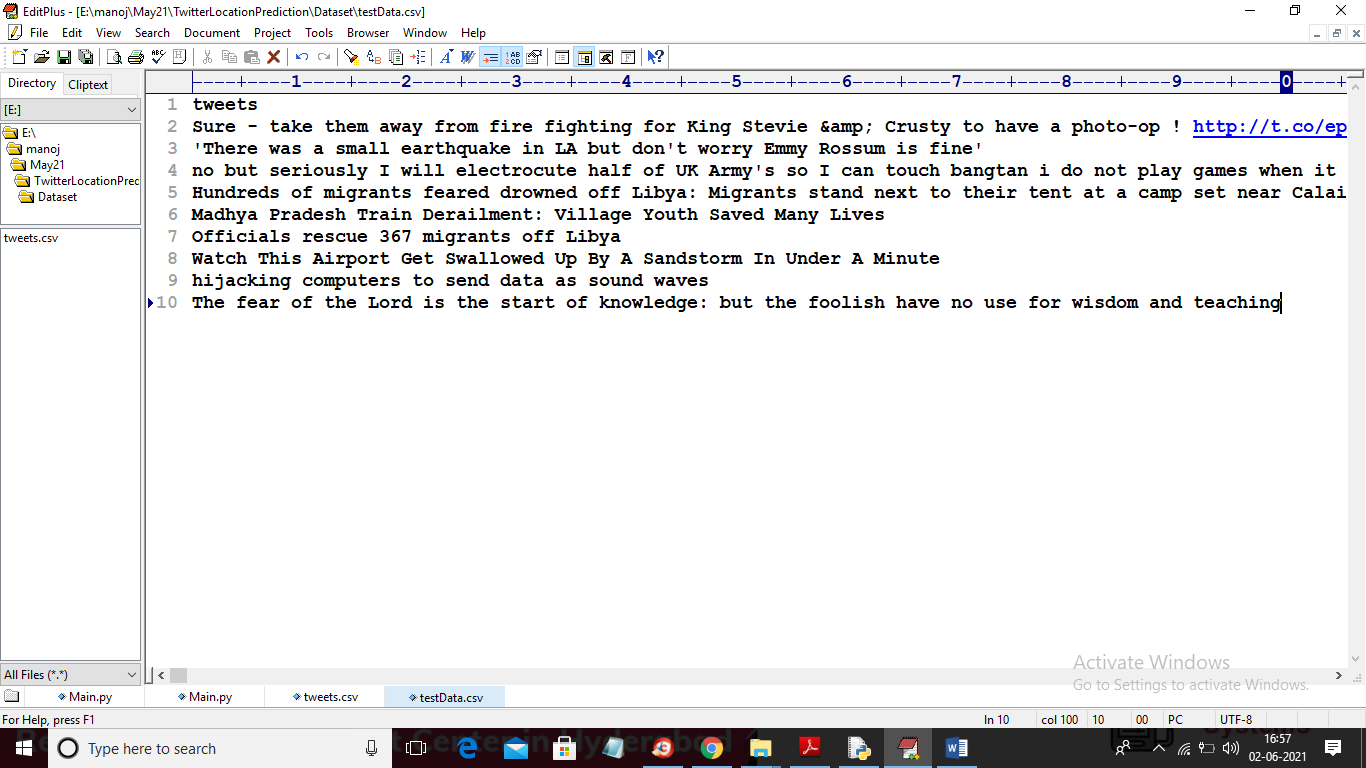
In this paper author is using machine learning algorithms such as SVM, Naïve Bayes and Decision Tree to predict location of twitter user by analysing his tweets. To predict location we have downloaded some tweets from twitter which consists of tweet messages and tweet location. We will train above machine learning algorithms with tweets dataset to predict user location. To implement this project we have designed following modules

1. Upload Dataset: using this module we will upload tweets dataset to application
2. Preprocess Dataset: tweets often contains raw data with special symbols stop words and URL’s and to make location prediction model we need to clean dataset. So dataset preprocess module will remove special symbols and stop words from dataset and make it clean. After cleaning data we will split dataset into train and test part where application used 80% dataset for training and 20% dataset to test trained model accuracy.
3. Run Machine Learning Algorithm: using this module we will train above 3 algorithms and then apply test data to check how may test records models are predicting correctly and based on that accuracy will be evaluated for each machine learning model.
4. Accuracy Comparison Graph: using this module we will accuracy comparison graph between all 3 algorithms
5. Predict Location from Test Tweets: using this module we will upload test tweets and then machine learning model will predict location of that test tweet.

We are using below dataset to train above machine learning algorithms and this dataset is saved inside ‘Dataset’ folder



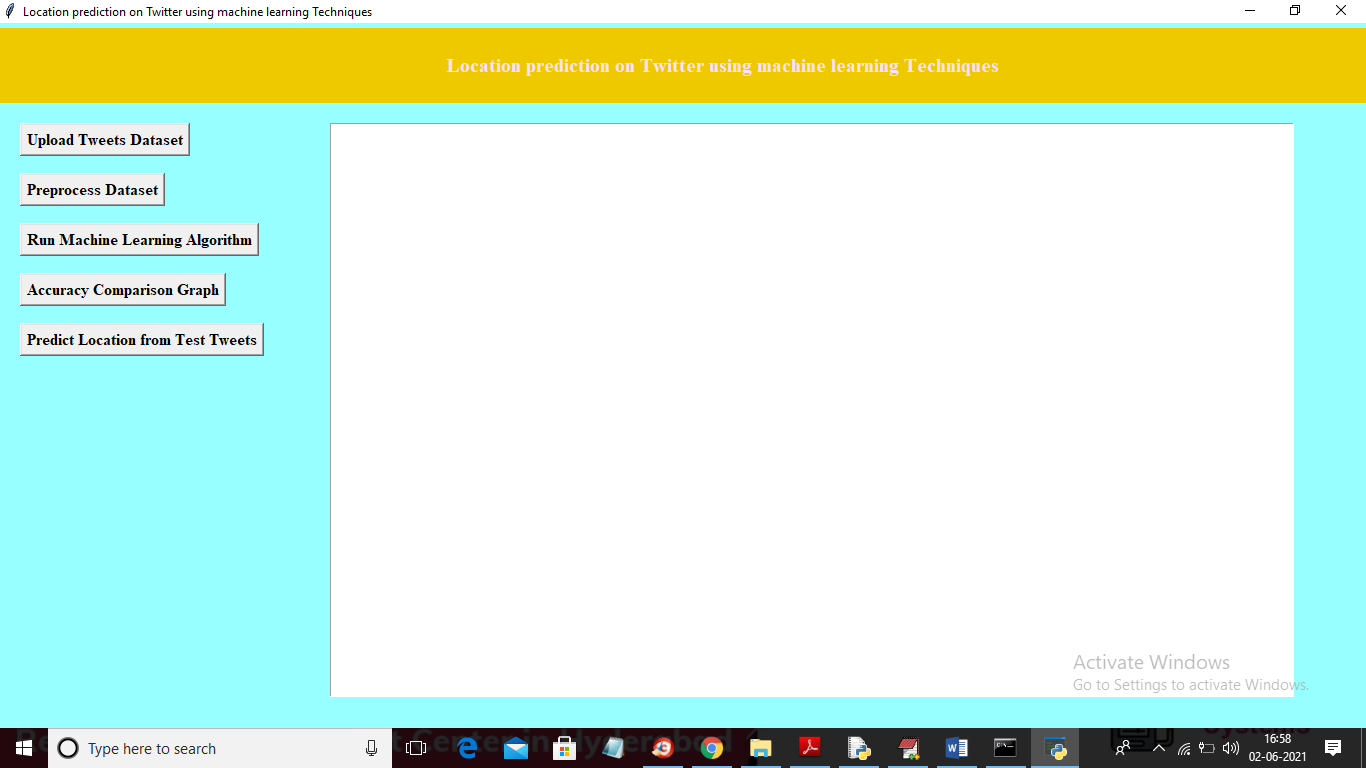
In above dataset screen first row contains dataset column names and remaining rows contains dataset values and in above dataset we are using ‘text and location’ column where text columns contains tweet message and location column contains tweet location. Below is the test dataset which contains only tweets and ML will predict location for that tweet



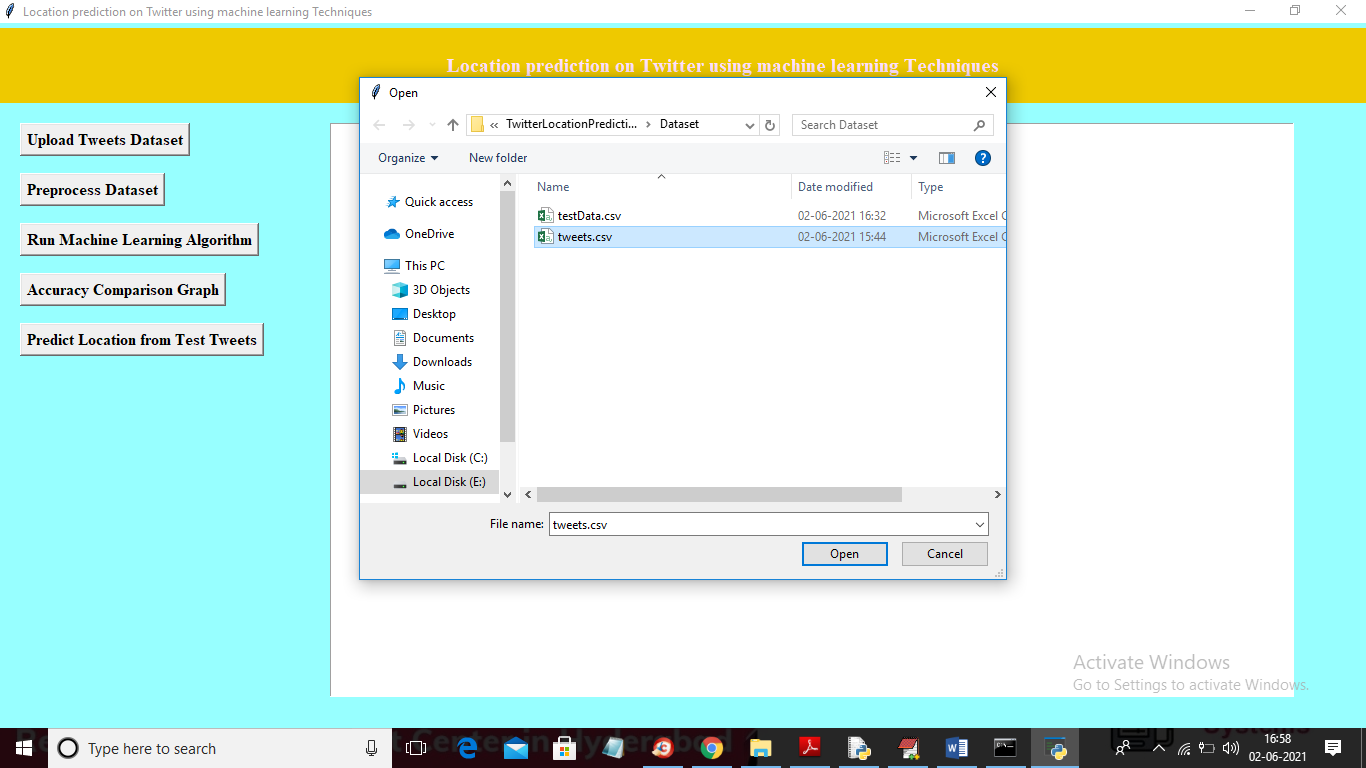
In above test dataset we can see only tweets messages are there and ML will predict location for each tweet

SCREEN SHOTS

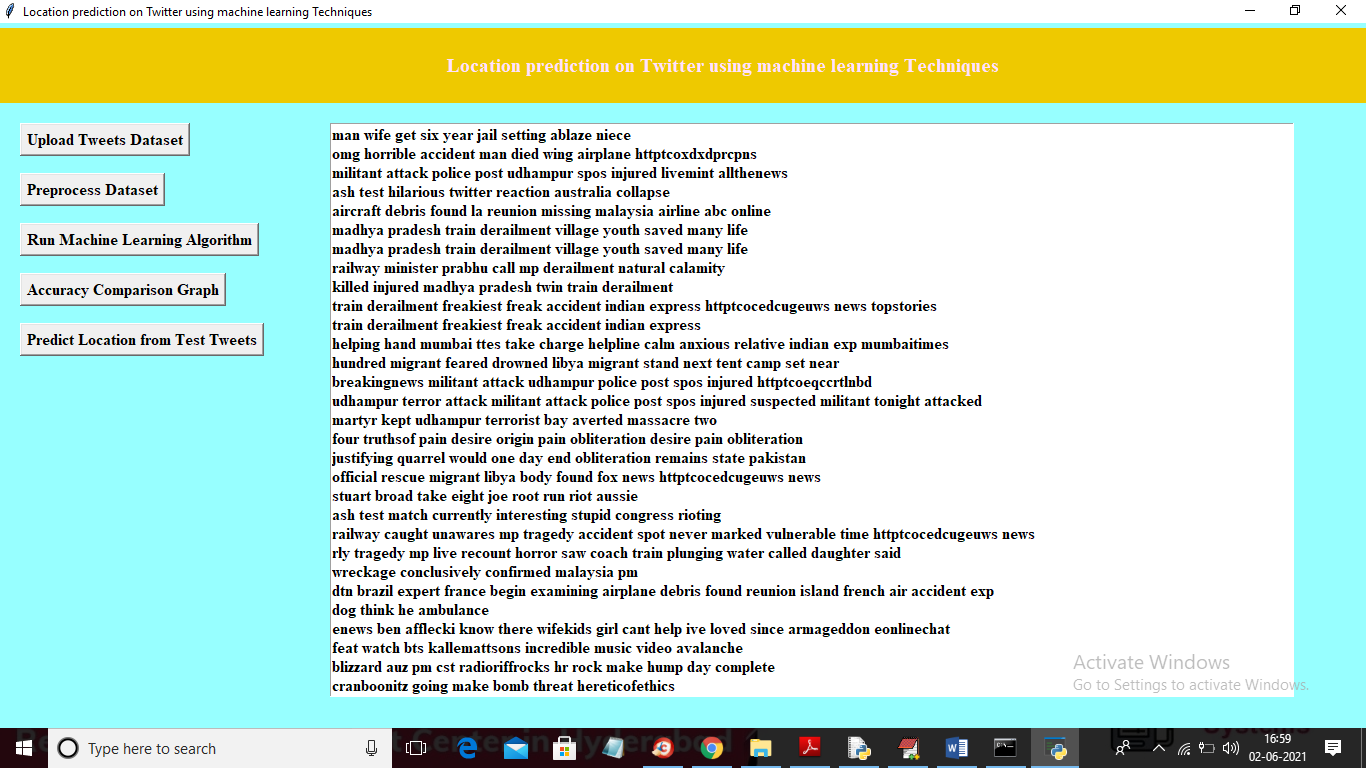
To run project double click on ‘run.bat’ file to get below screen



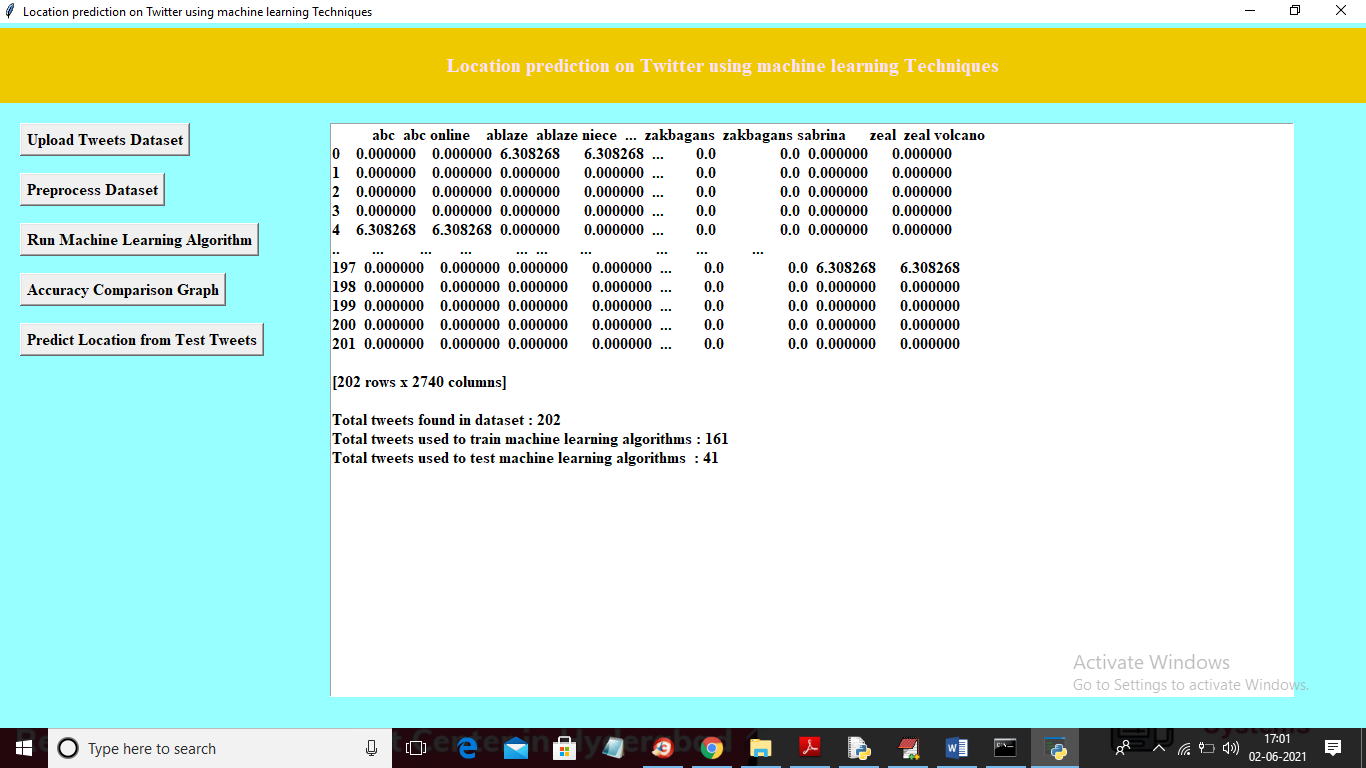
In above screen click on ‘Upload Tweets Dataset’ button to upload dataset



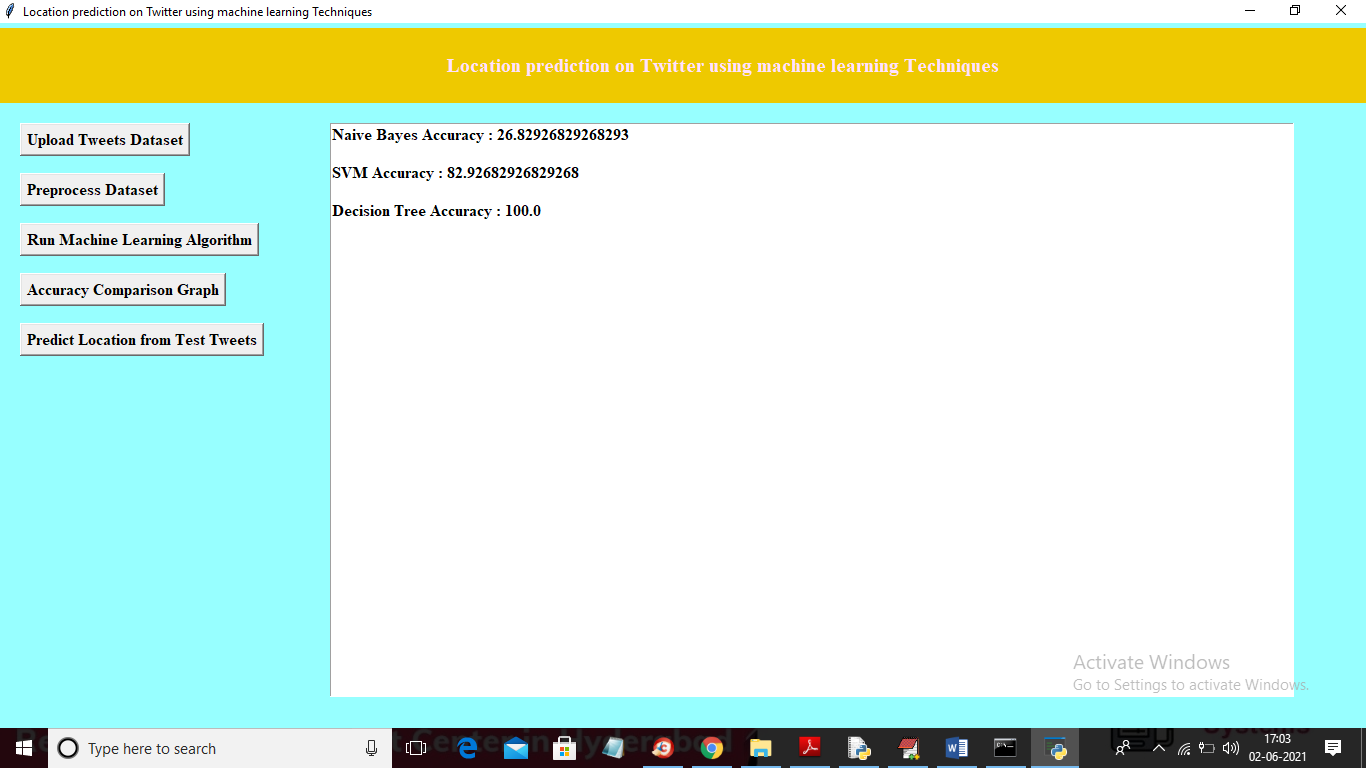
In above screen selecting and uploading ‘tweets.csv’ file and then click on ‘Open’ button to load dataset and to get below screen



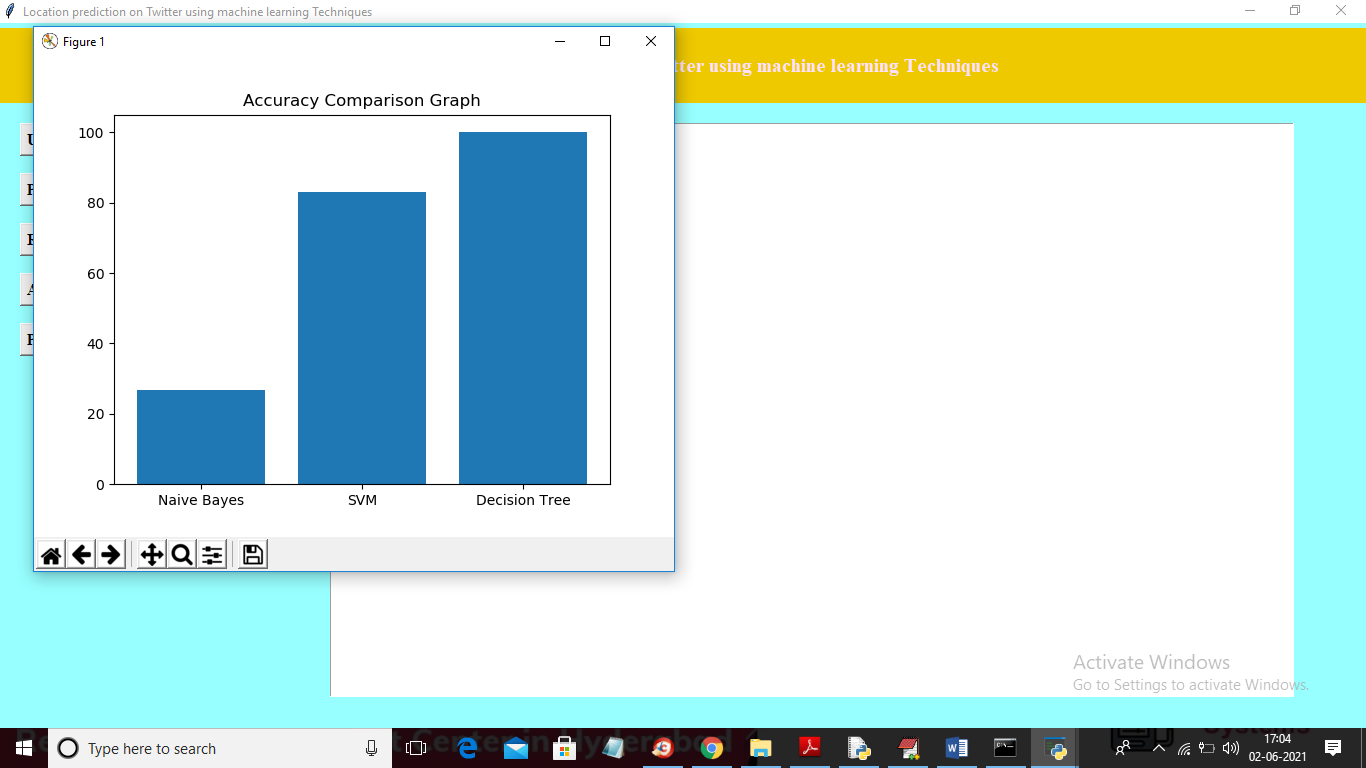
In above screen we can see all tweets are cleaned and displaying only tweets messages without special symbols and stop words and now click on ‘Preprocess Dataset’ to convert all tweets into TF-IDF vector and this vector will put all unique words into vector column and if word appear in tweet then that vector row will be filled with average of that words count and if word not appear then 0 will be put in vector column. So to convert tweets into vector click on ‘Preprocess Dataset’ button



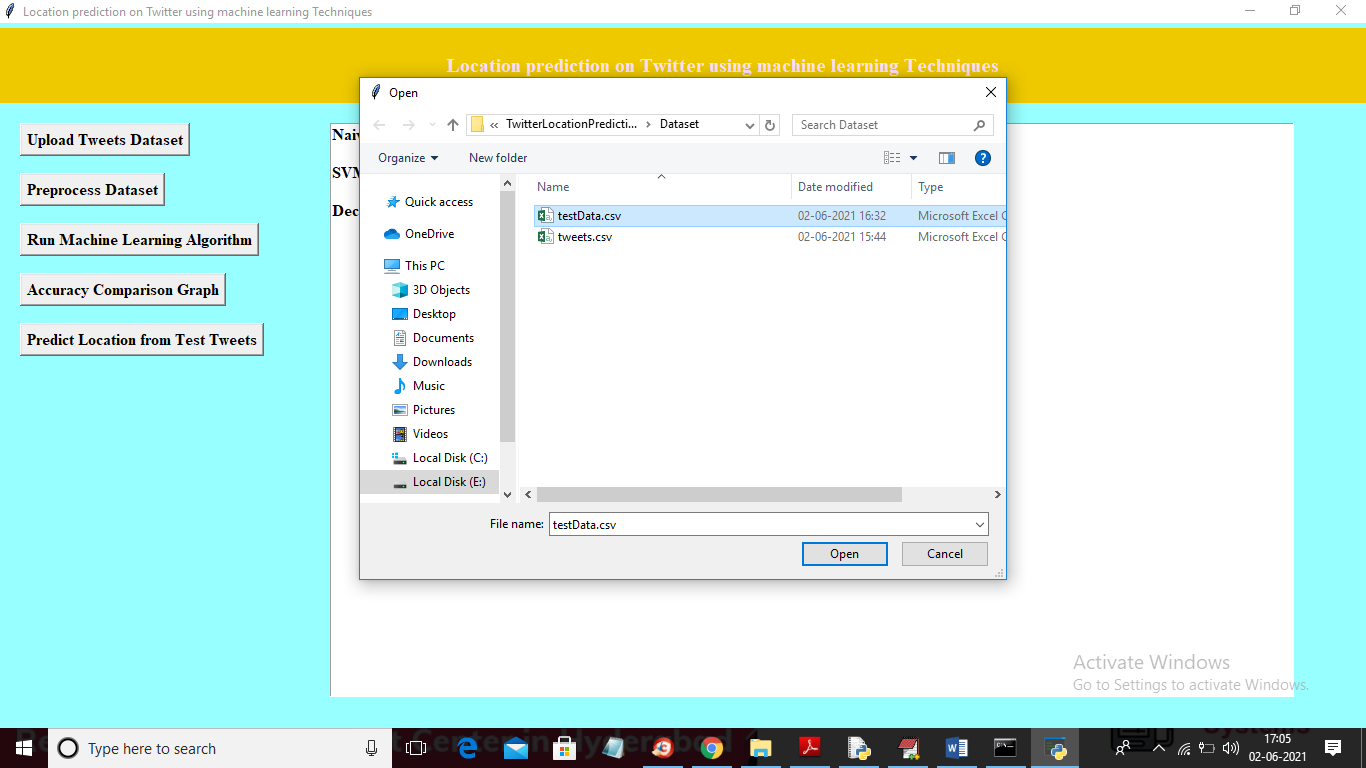
In above screen we can see in first row we have words in vector and remaining rows showing count average of that word. Now vector is ready and now click on ‘Run Machine Learning Algorithms’ button to train all 3 ML with above vector and calculate accuracy



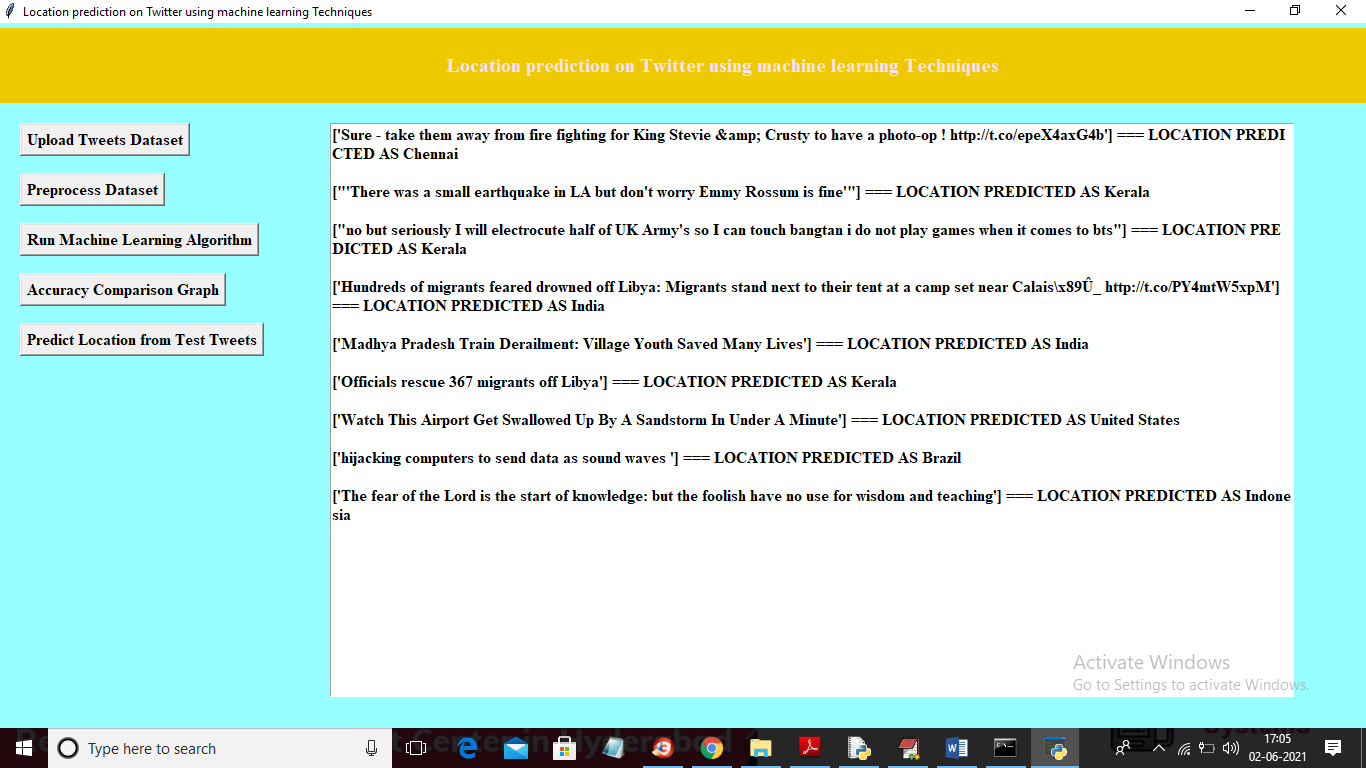
In above screen Naïve Bayes got 26% accuracy and SVM got 82 and Decision Tree got 100% accuracy and now ML models are ready and now click on ‘Accuracy Comparison Graph’ button to get below graph



In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and now click on ‘Predict Location from Test Tweets’ button to upload test tweets and then ML will predict location



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



In above screen in square bracket we can see tweet message and after square bracket we can see predicted location for that tweet. In above screen first tweet message location predicted as ‘Chennai'