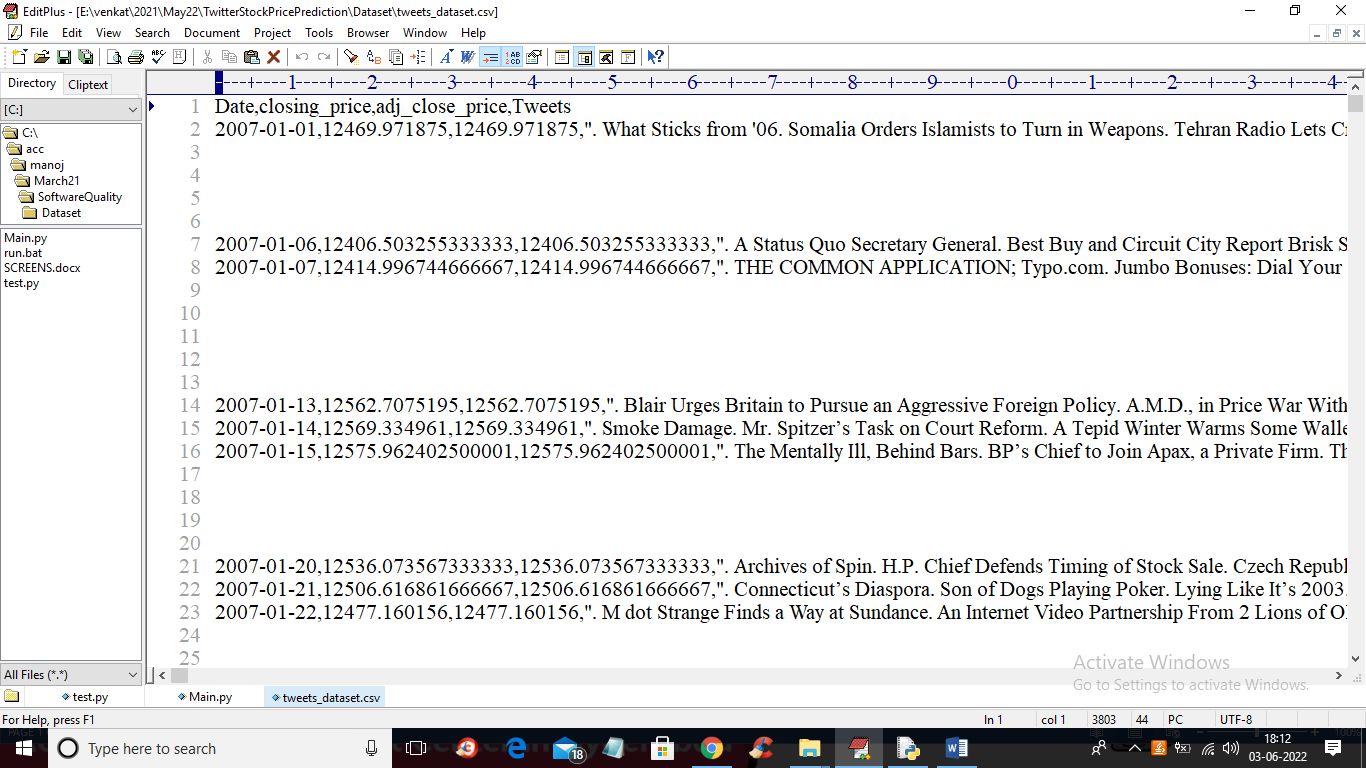
Stock Price Prediction using Twitter Dataset

In this project we are using twitter tweets dataset to predict stock prices as customers will use social media to describe their share market experience and by using their experiences we can predict stock prices. If stock performing well then customers will gain from it and give positive reviews or TWEETS and if not performing well then then will give negative tweets and by using such experiences we can train machine learning algorithms to build stock price prediction model and when we input new test data then this model will predict stock prices based on given test data and future values.

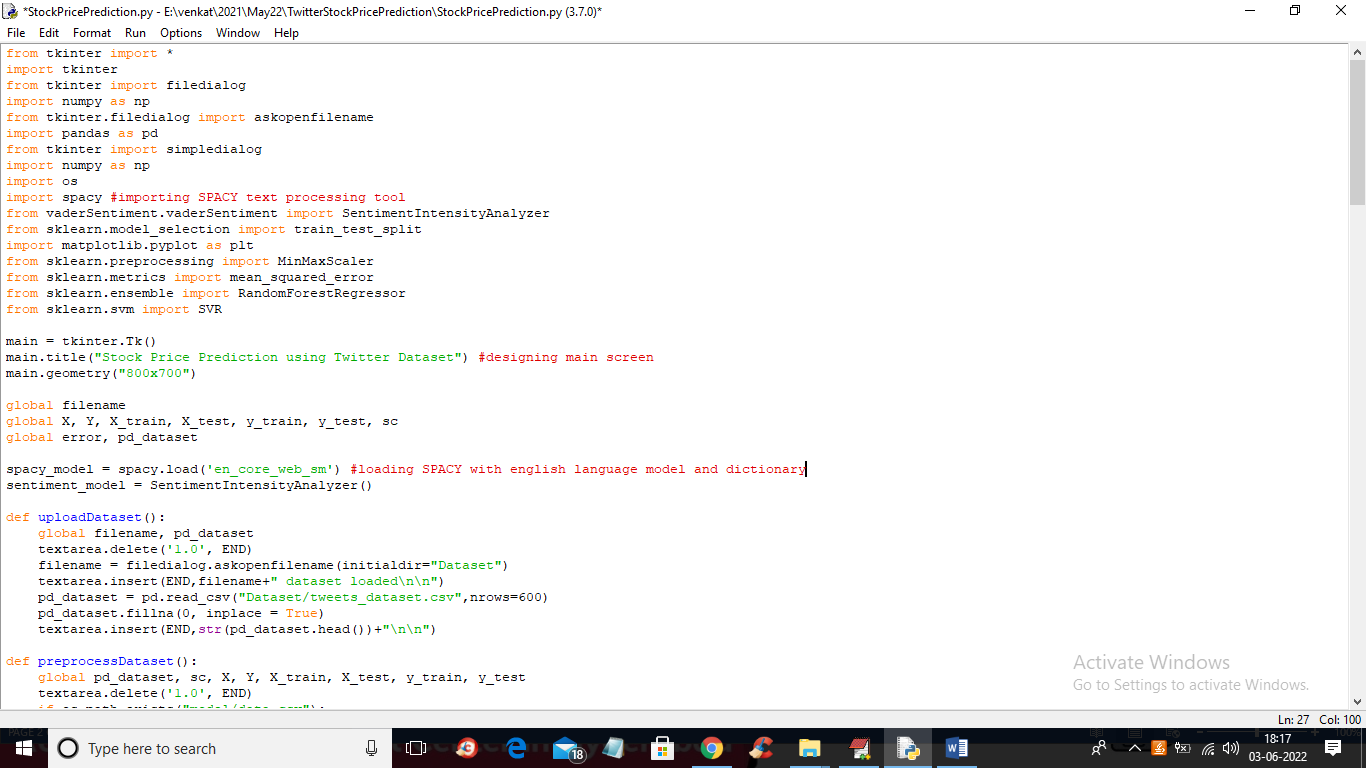
In propose work we are using SVM and Random Forest algorithms to train stock price prediction model and we are using below tweets dataset as input to those algorithms

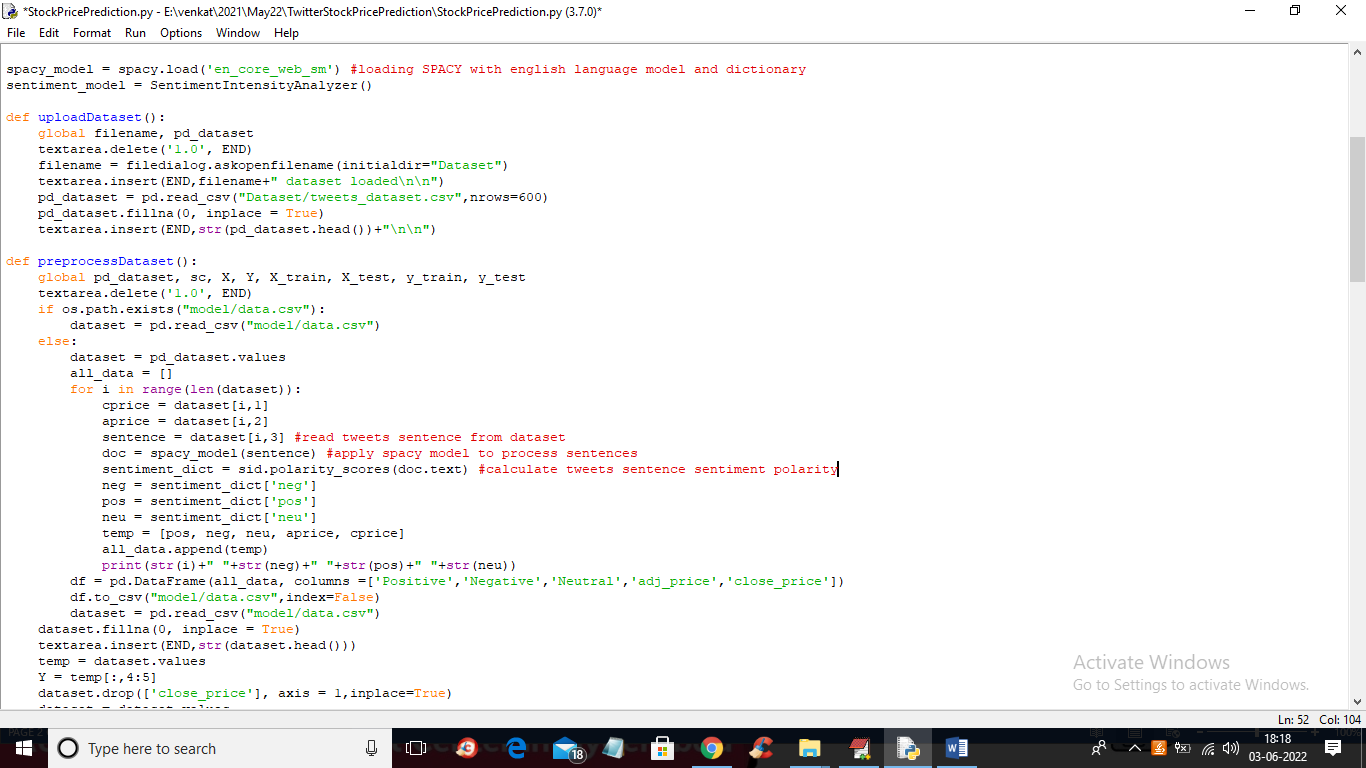


In above dataset first row contains dataset columns names and remaining rows contains dataset values as stock prices TWEETS and we will input this to ML algorithms.

All ML algorithms will take only numeric values as input but tweets contains non-numeric data so we will be applying SPACY text processing tool to extract tweets sentences and then remove special symbols and stop words and then find probability of tweets sentiments as POSITIVE, NEGATIVE and NEUTRAL and if customer happy then tweets will contains more positive probability and by using this numeric probabilities we will train ML algorithms.

In below screen we are using SPACY text processing tool to process tweets





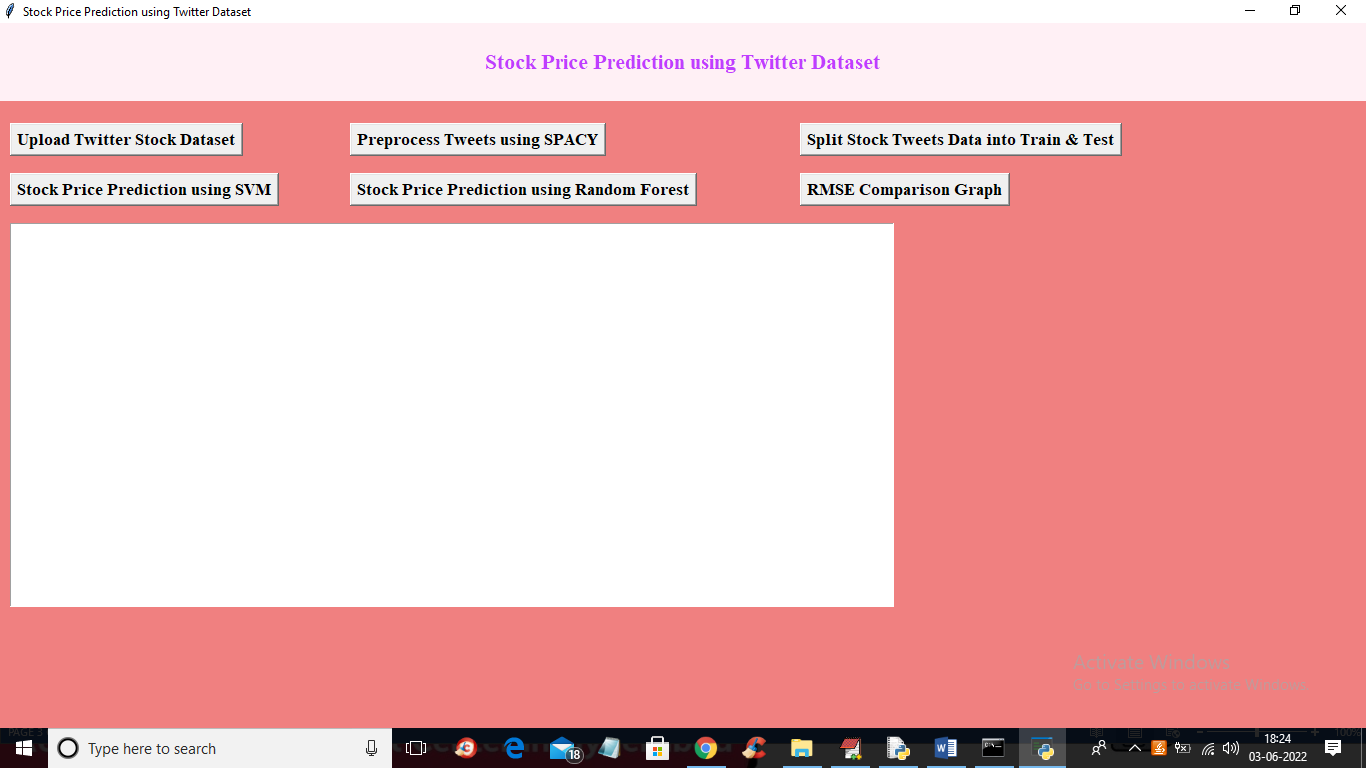
In above 2 screens read red colour comments to know about SPACY implementation

Modules Information

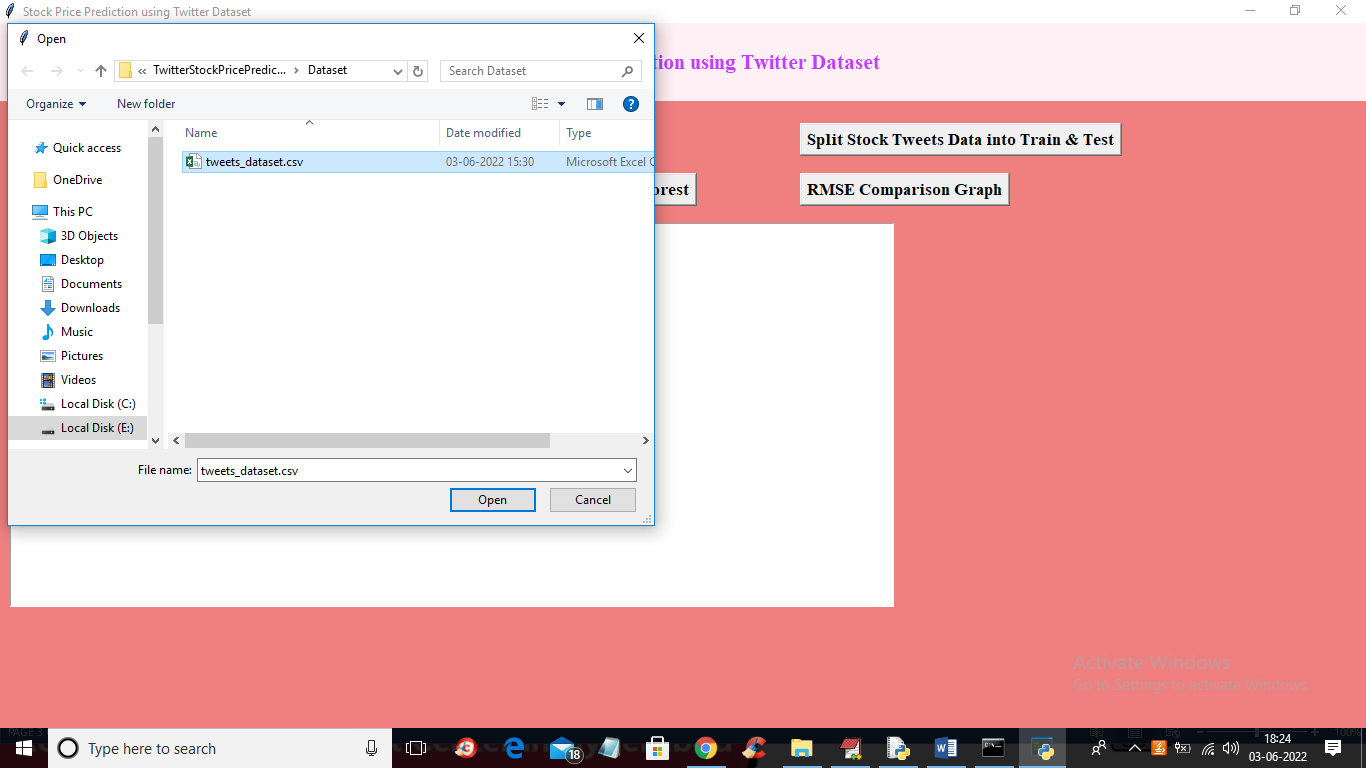
1. Upload Twitter Stock Dataset: using this module we will upload tweets dataset to application
2. Preprocess Tweets using SPACY: using this module we will apply SPACY tool to process tweets to extract sentiments and then convert dataset into numeric sentiment positive, negative and neutral probability
3. Split Stock Tweets Data into Train & Test: using this module we will split dataset values into train and test where application used 80% dataset for training and 20% for testing
4. Stock Price Prediction using SVM: using this module we will train SVM on 80% dataset and then build price prediction model. 20% dataset will be applied on this model to predict prices and then calculate difference between original test data and predicted price data as error or ROOT MEAN SQUARE ERROR (RMSE). The lower the error the better or accurate is the stock price prediction
5. Stock Price Prediction using Random Forest: using this module we will train Random Forest on 80% dataset and then build price prediction model. 20% dataset will be applied on this model to predict prices and then calculate difference between original test data and predicted price data as error or ROOT MEAN SQUARE ERROR (RMSE). The lower the error the better or accurate is the stock price prediction
6. RMSE Comparison Graph: using this module we will plot error graph between SVM and Random Forest

SCREEN SHOTS

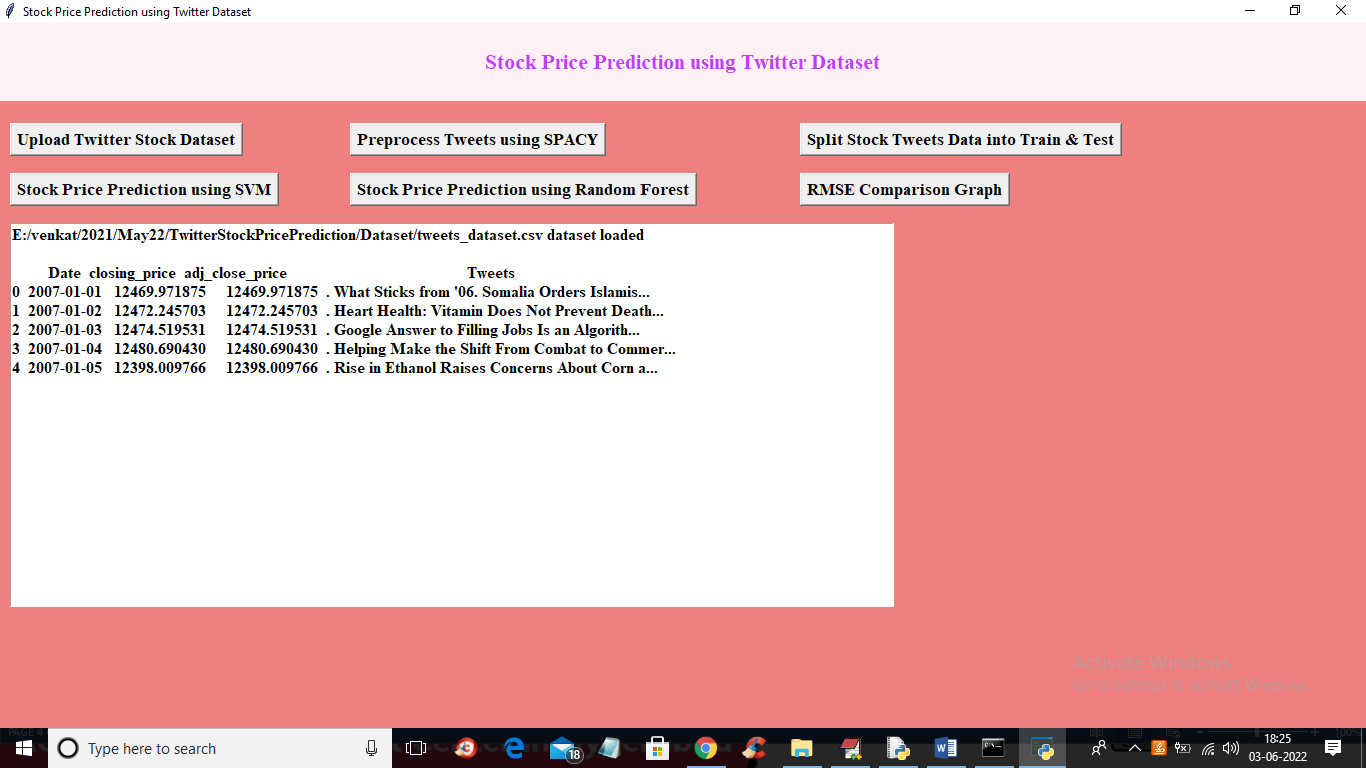
To run project double click on ‘run.bat’ file to get below screen



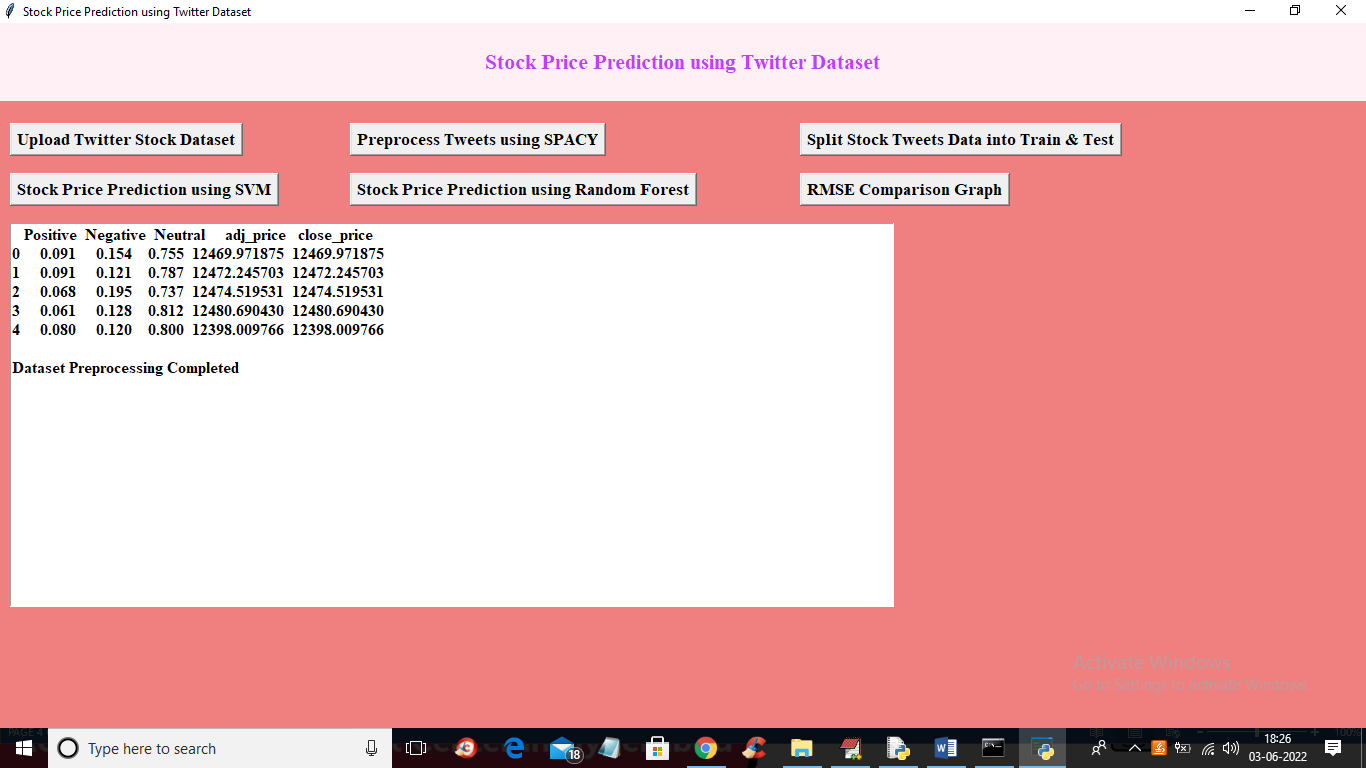
In above screen click on ‘Upload Twitter Stock Dataset’ button to upload dataset and get below screen



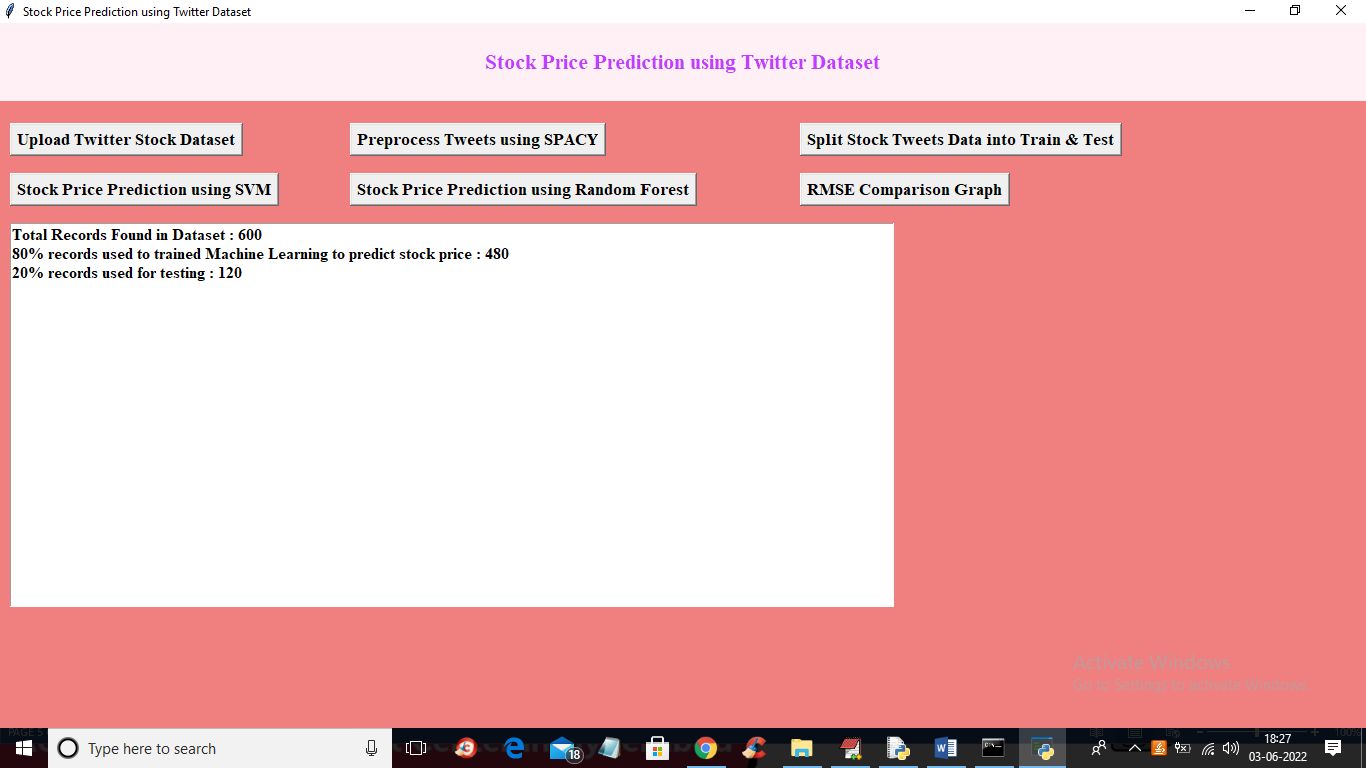
In above screen selecting and uploading tweets stock price dataset and then click on ‘Open’ button to load dataset and get below output



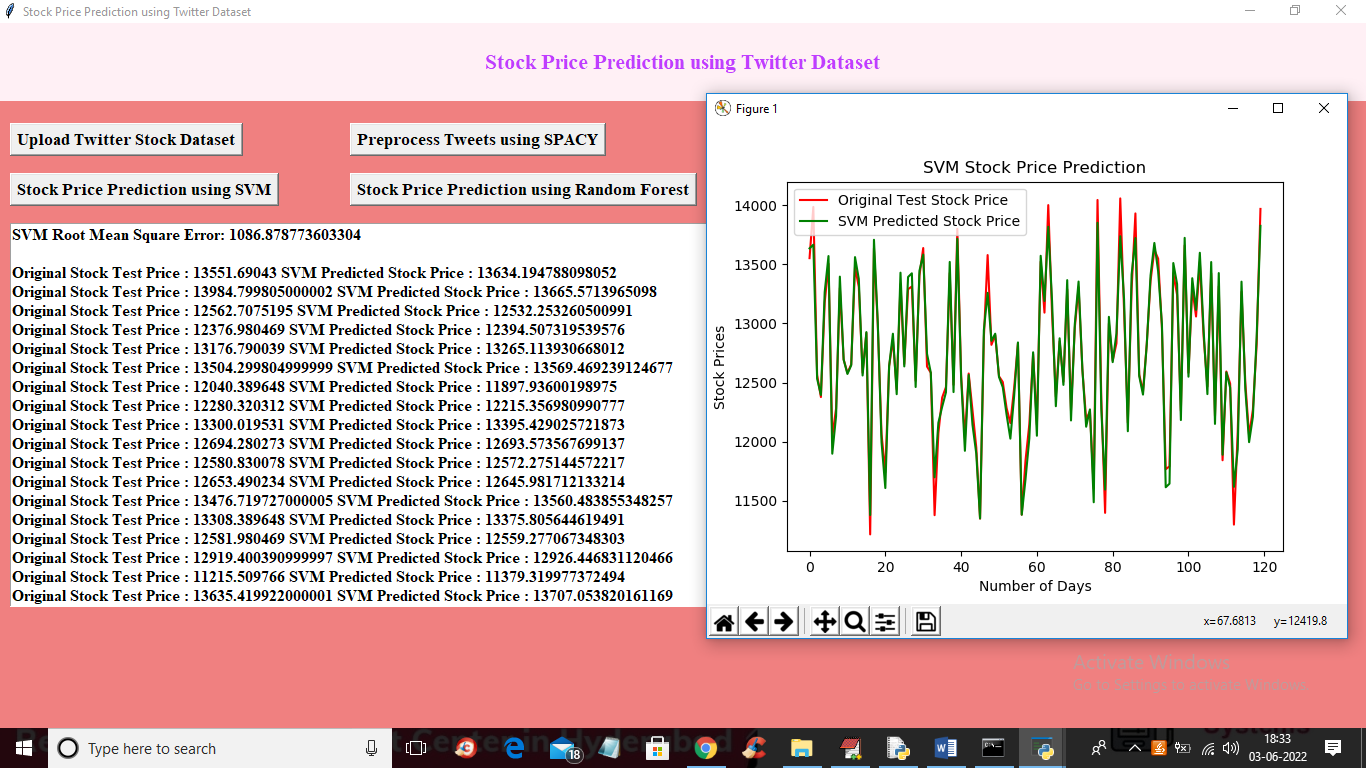
In above screen tweets and stock prices loaded and now click on ‘Preprocess Tweets using SPACY’ button to read all tweets and then find sentiment probability of each tweet in terms of positive, negative and neutral and then will get below output



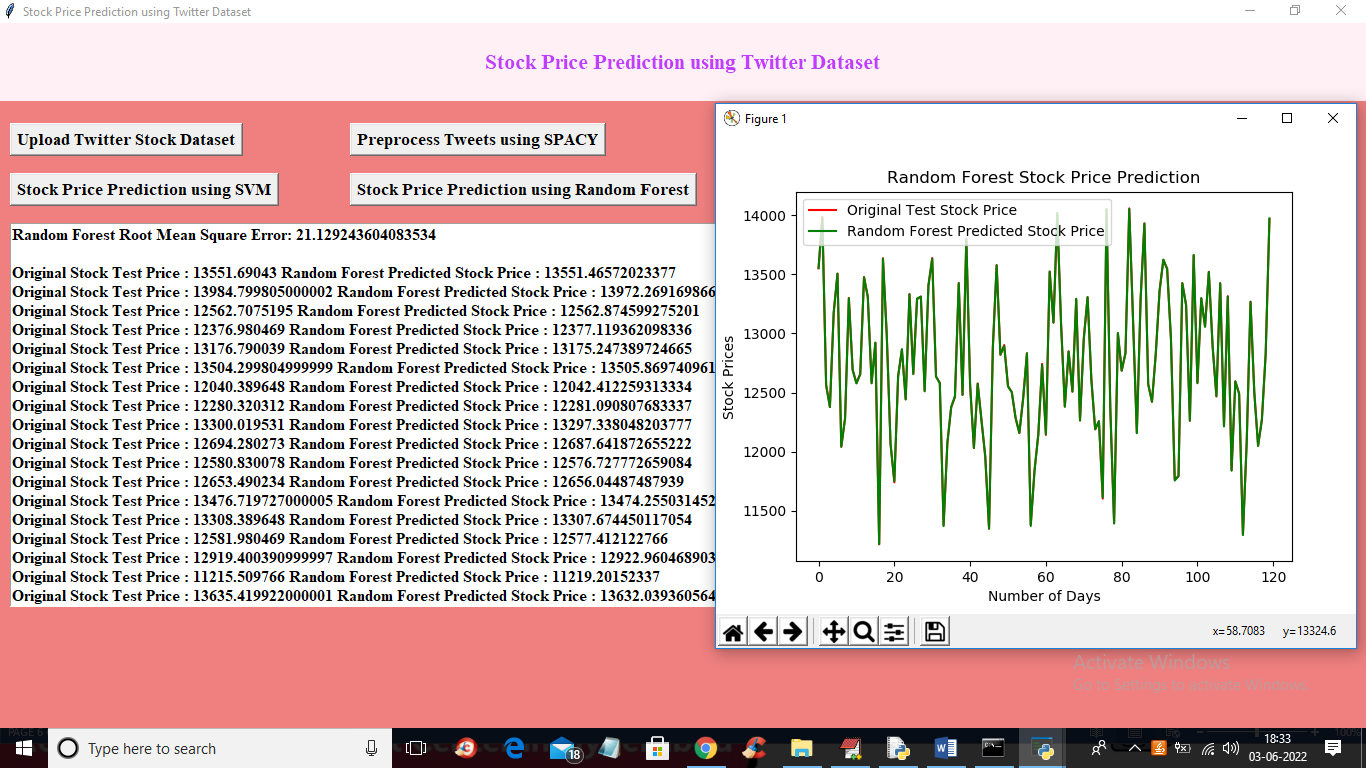
In above screen for tweet we got customer experience in terms of sentiments of stock performance and now click on ‘Split Stock Tweets Data into Train & Test’



In above screen we can see dataset contain 600 tweets and application using 80% (480) tweets for training and 20% (120) tweets for testing and now train and test data is ready and now click on ‘Stock Price Prediction using SVM’ button to train SVM and get below output

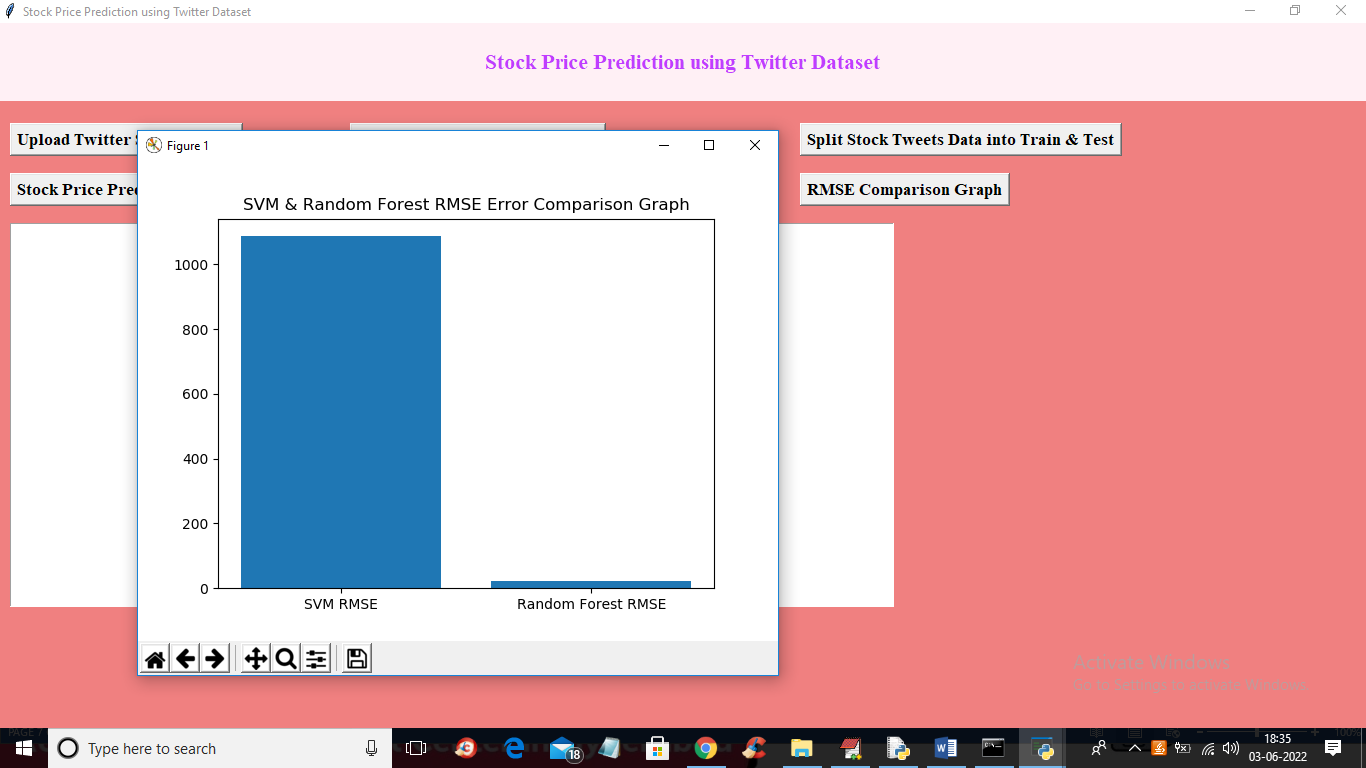


In above screen in text area we can see original test stock prices and SVM stock prices and we got SVM RMSE as 1086(difference between predicted and original prices) and in above graph x-axis contains NUMBER of DAYS and y-axis represents STOCK prices and red line represents original stock prices and green line represents SVM predicted prices and we can see both lines are overlapping with little gap and difference so we can say SVM performance is good but not best so close above graph and then click on ‘Stock Price Prediction using Random Forest’ button to train with Random Forest and get below output



In above screen with Random Forest we got RSME as just 21 and in graph we can see both lines are fully overlapping so TEST prices and Random Forest predicted prices are accurate. So Random Forest is best in performance

Now close above graph and then click on ‘RMSE Comparison Graph’ button to get below graph



In above graph x-axis represents algorithm names and y-axis represents error rate and in both algorithms Random Forest got less error so its performance is best. The lower the error the better is the algorithm