Spammer Detection and Fake User Identification on Social Networks

In this paper author is describing concept to detect spam tweets and fake user account from online social network called twitter. To perform detection author is using twitter dataset and 4 different techniques called Fake Content, Spam URL Detection, Spam Trending Topic and Fake User Identification. Using above 4 techniques we can identify whether tweet is normal or spam and then using Random Forest data Mining algorithm we will train above dataset to classify number of spam and non-spam tweets or fake or non-fake accounts. For each technique author is using different data mining techniques to classify tweets as spam or non-spam but here we are using Random Forest classifier.

Description of 4 techniques to detect tweet is spam or normal.

The presented techniques are also compared based on various features, such as user features (retweets, tweets, followers etc.), content features (tweet content messages).

1. Fake Content: If the number of followers is low in comparison with the number of followings, the credibility of an account is low and the possibility that the account is spam is relatively high. Likewise, feature based on content includes tweets reputation, HTTP links, mentions and replies, and trending topics. For the time feature, if many tweets are sent by a user account in a certain time interval, then it is a spam account.
2. Spam URL Detection: The user-based features are identified through various objects such as account age and number of user favourites, lists, and tweets. The identified user-based features are parsed from the JSON structure. On the other hand, the tweet-based features include the number of (i) retweets, (ii) hashtags, (iii) user mentions, and (iv) URLs. Using machine learning algorithm called Naïve Bayes we will check whether tweets contains spam URL or not.
3. Detecting Spam in Trending Topic: In this technique tweets content will be classified using Naïve Bayes algorithm to check whether tweet contains spam or non-spam words. This algorithm will check for spam URL, adult content words and duplicate tweets. If Naïve Bayes detect tweet as SPAM then it will return 1 and if not detected any SPAM content then Naïve Bayes will return 0.
4. Fake User Identification: These attributes include the number of followers and following, account age etc. Alternatively, content features are linked to the tweets that are posted by users as spam bots that post a huge amount of duplicate contents as contrast to non-spammers who do not post duplicate tweets. In this technique features (following, followers, tweet contents to detect spam or non-spam content using Naïve Bayes Algorithm) will be extracted from tweets and then classify those features with Naïve Bayes Algorithm as spam or non-spam. Later this features will be train with random forest algorithm to determine account is fake or non-fake. All extracted features will be saved inside features.txt file. Naïve Bayes classifier saved inside ‘model’ folder.

Using above techniques we can detect whether tweets contains normal message or spam message. By detecting and removing such spam messages help social networks in gaining good reputation in the market. If social networks did not remove spam messages then its popularity will be decreases. Now a days all users are heavily dependent on social networks to get current news and business and relatives information and thus protecting it from spammer help it to gain reputation.

To implement this project we are using twitter dataset in JSON format and this dataset contains user details, tweets count, follower, following, favourites, tweet text etc. Using Python JSON API we reading all details to detect whether user account is fake or genuine or contains spam or normal messages. All this dataset files are available inside ‘tweets’ folder.

Technologies used to implement this project

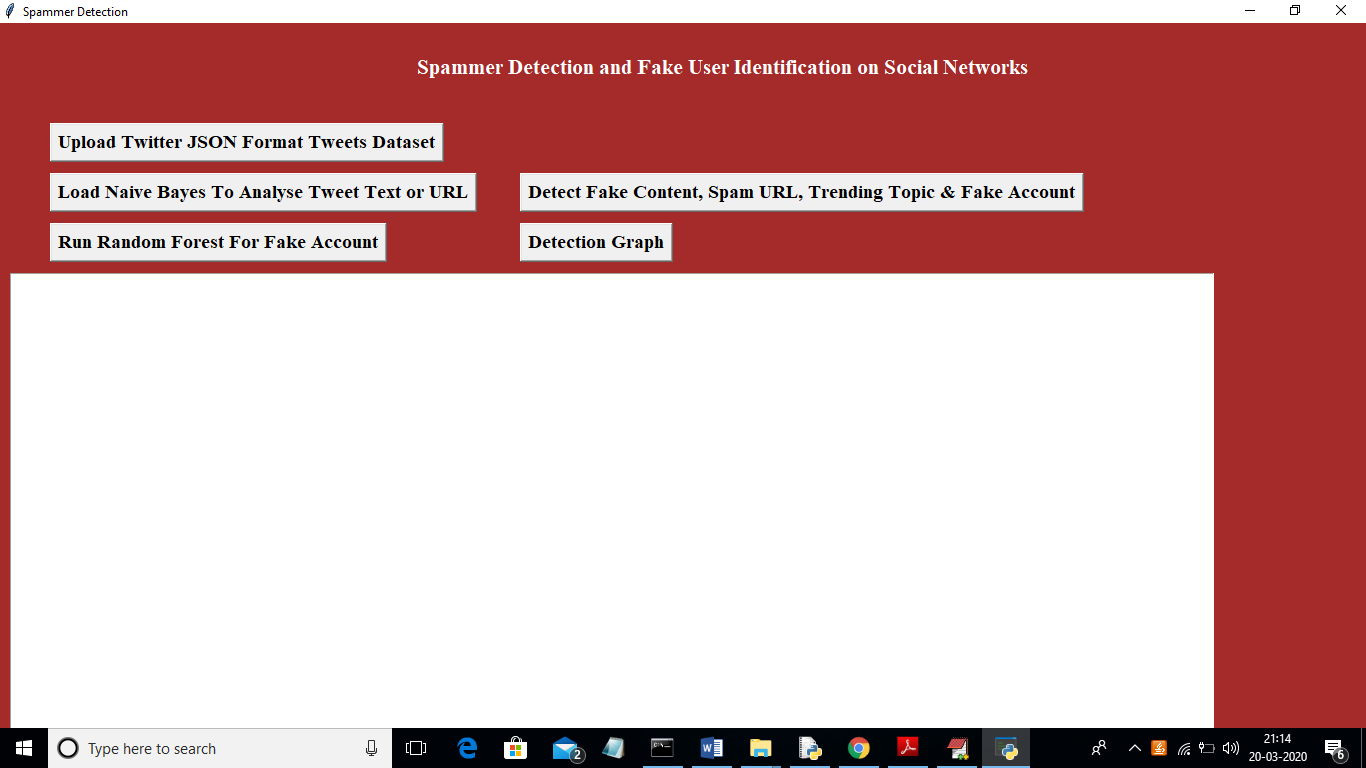
Programming Language: Python

Packages : SKlearn and Numpy

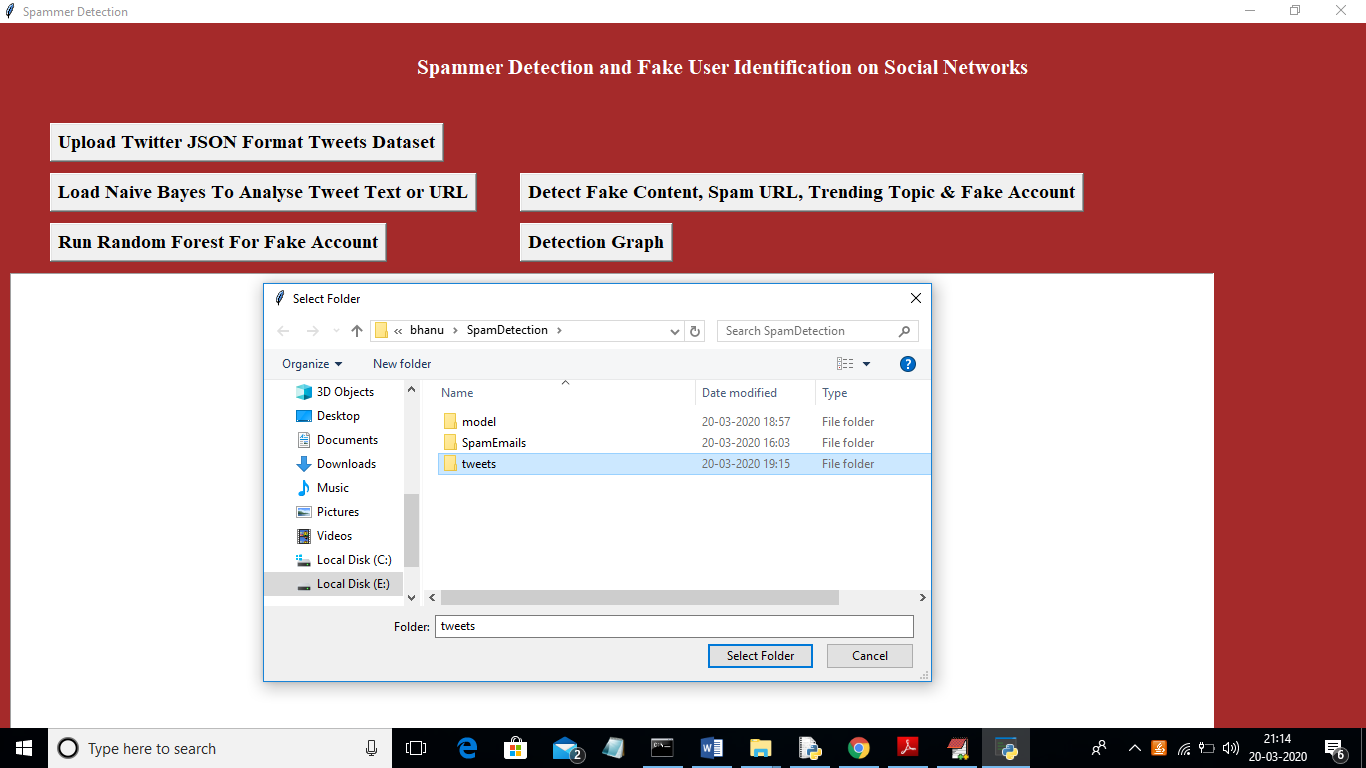
GUI : Python TKINTER

Screen shots

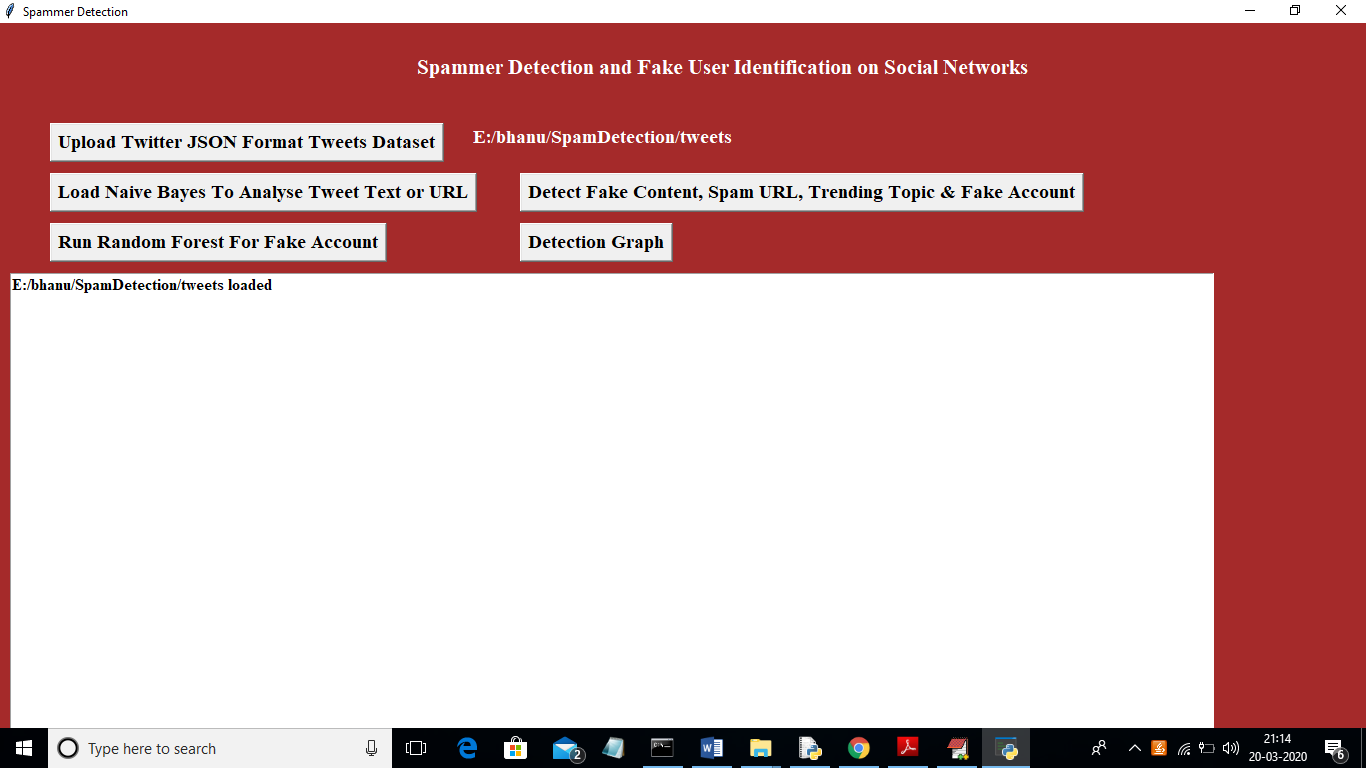
To run this project double click on ‘run.bat’ file to get below screen



In above screen click on ‘Upload Twitter JSON Format Tweets Dataset’ button and upload tweets folder



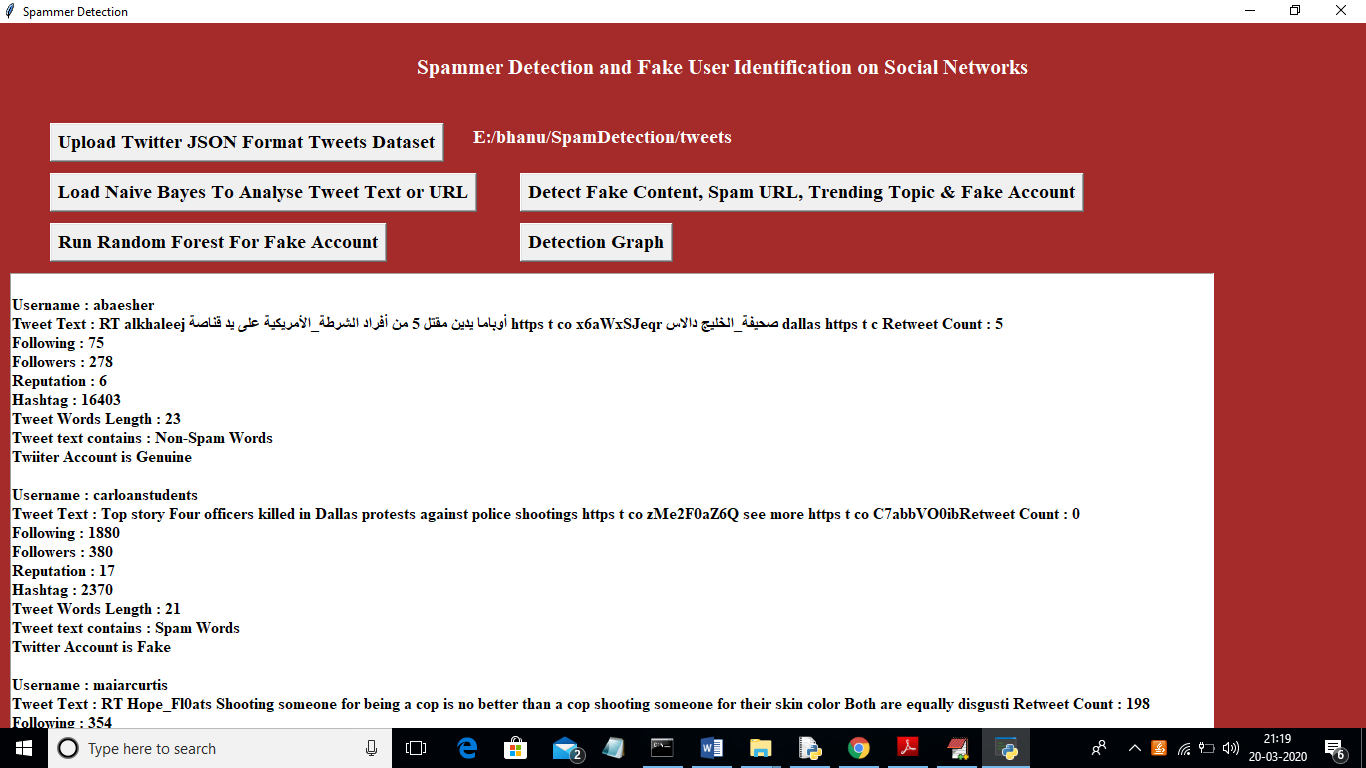
In above screen I am uploading ‘tweets’ folder which contains tweets from various users in JSON format. Now click open button to start reading tweets



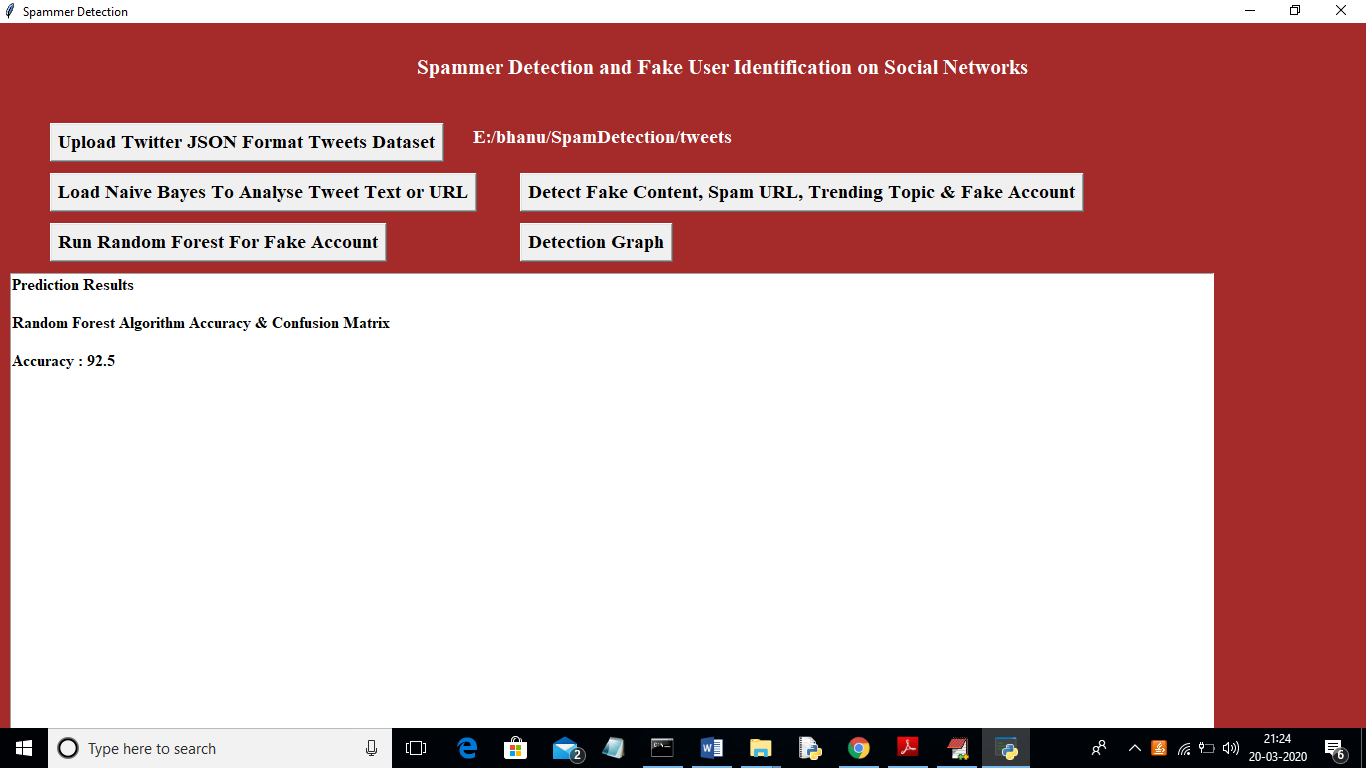
In above screen we can see all tweets from all users loaded. Now click on ‘Load Naive Bayes To Analyse Tweet Text or URL’ button to load Naïve Bayes classifier



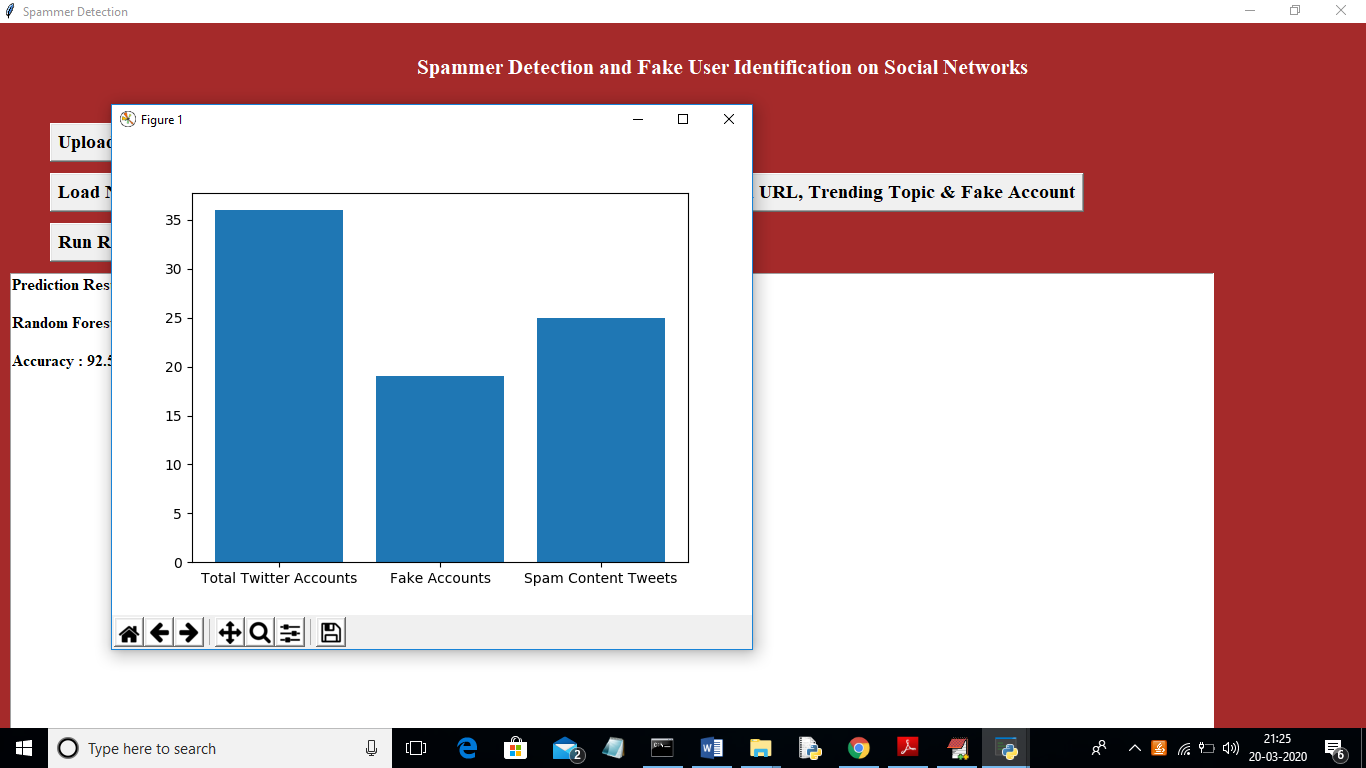
In above screen naïve bayes classifier loaded and now click on ‘Detect Fake Content, Spam URL, Trending Topic & Fake Account’ to analyse each tweet for fake content, spam URL and fake account using Naïve Bayes classifier and other above mention technique



In above screen all features extracted from tweets dataset and then analyse those features to identify tweets is no spam or spam. In above text area each records values are separated with empty line and each tweet record display values as TWEET TEXT, FOLLOWERS, FOLLOWING etc with account is fake or genuine and tweet text contains spam or non-spam words. Now click on ‘Run Random Forest Prediction’ button to train random forest classifier with extracted tweets features and this random forest classifier model will be used to predict/detect fake or spam account for upcoming future tweets. Scroll down above text area to view details of each tweet



In above screen we got random forest prediction accuracy as 92%, now click on ‘Detection Graph’ button to know total tweets and spam and fake account graph



In above graph x-axis represents total tweets, fake account and spam words content tweets and y-axis represents count of them