Documentation

Twitter User Analysis

Submitted By:

IIT2016106 - Aswin VB

IIT2016105 - Aswanth K

IIT2016133 - Utpal Aman

BIM2016501 - Druval CR

IRM2016006 - Sarath Nandimandalam

Directory Structure

- Files related to algorithm used is present in subfolder Twitter folder
- Main.py in Twitter folder contains the algorithms which is to be run from which other sub functions are called
- Dataset_get.csv is the dataset for a,b,c,d,e,FAL values
- Classifier.py takes in the data present in the dataset_gen.csv and classify the users based on different Classification Algorithm
- wot.py is used to calculate Web Of Trust Rank
- similarity.py is used to calculate similarity of tweets
- url.py is used to calculate Alexa rank of url's present in the tweets
- checkTime.py is used to calculate time difference of tweets
- checkContent.py is used to check for adult contents in tweets
- Main directory files are used for the django application

How To Run?

To check whether a particular user is anomalous (using webapp)

- Open the folder in terminal and run the following commands
- virtualenv venv
- source venv/bin/activate
- pip install -r requirements.txt
- python manage.py migrate
- python manage.py runserver
- open localhost:8000/main in your browser
- Enter the username which you want to test
- Output will be displayed after sometime

To do the classification/check particular user(in command line) follow these steps

- open the subfolder named Twitter folder in terminal
- Run main.py
- You will get two option
  - Check anomaly for a particular user (1)
  - Classify a dataset of twitter usernames (2)

- If the selected option is 1 then you will be prompted to enter the username and on entering the username the algorithm will find out the status of the user
- If the selected option is 2 then you will get two suboptions
  - Generate dataset of values a,b,c,d,e and then classify (Y)
  - Classify pre built dataset (N)
- If the selected option is “Y” then the usernames present in “Followers.csv” is applied onto the algorithm and a,b,c,d,e,FAL values are found and saved in “dataset_gen.csv” . Once it is completed the dataset is applied onto the classification algorithms and its confusion matrix along with its accuracy will be displayed
- If the selected option is “N” then the prebuilt data in the “dataset_gen.csv” is used and classified and its confusion matrix along with its accuracy will be displayed