

Approach:

We have collected the tweets for the given subject based on the name and keywords given. And then we have analyzed multiple content like if the subject's is bully or not, hate speech, attacking nature, sexist or not and predicted the personality traits based on the BIG-5 model.

Output:

Output of the code will be two files, namely 'Information.txt' and 'Predictions.txt'. Information.txt file contains the tweets tweeted by the subject. Predictions.txt file contains the predictions that are predicted as mentioned above.

Instructions to run the code:

1. Make sure that all the files extracted are in the same folder.
2. Make sure that your laptop/PC have an active internet connection.
3. 'pip' should be installed in your laptop/PC.
4. Create a virtual environment by using-
 - a. `virtualenv -p python3.8 <env-name>` (`<env-name>` -- replace with environment name)
5. If virtualenv is not installed please install it by
 - a. `pip install virtualenv`
6. Then activate the environment by using-
 - a. `<env name>\Scripts\activate.bat`
 - b. Then run the code by command line.
7. Our code will install the packages required automatically, if unfortunately, any of the packages are not installed automatically, then please install the corresponding packages manually. (`pip install <package-name>`)
8. During the run time if any of the packages are not installed then there will be some output regarding the installation process.
9. Information.txt: This file contains the information collected for the given details of the subject in the input file and are separated by "TESTCASE" and details of the subject.
10. Predictions.txt: This file contains the predictions that are made based on the information collected using the details of the subject collected from the input file. The predictions of each person are separated by "TESTCASE" and details of the subject.

Note: We have attached two codes (I) MainCode.py (ii) LighterCode.py

MainCode.py is responding slowly due to slow response of google translator API and BERT model in personality classification. On an average it takes 5 minutes to predict each subject's personality.

LighterCode.py doesn't contain the translator feature and BERT model, so it gives the output 10 times faster than the MainCode.py file.

So, depending on time and computational power use the correct file.

Try to use the MainCode.py as it gives more accurate predictions.