

HW1 – Sentiment Analysis on Twitter

This homework assignment is to do the sentiment analysis based on a text file which contains numerous tweets in its raw text format. All tweets collected in this file contains a key word specified by the instructor. The key word being used is “Trump”.

You can use the raw tweets file (*Trump_Raw_Tweets.txt*) provided by instructor directly or you can generate your own raw tweets file by following instructions in the file “*Twitter_Setup_V2.pptx*”. Your raw tweets file size at least needs to be 350K.

After you have the raw tweets file ready, you need to perform the following tasks in a Jupyter notebook file.

First you need to clean up the tweet file content to the best you can. For each word that is not a “stop word” (and, the, a, is, as, ...), assign a value +1 for positive sentiment or a value -1 for negative sentiment. A list of “positive” words will be provided to you. You can easily find a list of “stop words”, and a list of “negative” words online. For the words that is not in positive/negative/stop words, count as ‘others’. You need to answer the following questions:

- What’s the word count for positive/negative/stop word/others?
- What’s the ratio of positive/negative/stop word/others compare to the total word count?
- What’s the ratio for positive vs negative word count?
- Do you think that the general sentiment is negative or positive? Weakly or strongly?

Submission

Submit one Jupyter notebook file (.ipynb) with your code; be sure to explain your code using mark downs cell and comments. No need to attach any other file.