

Pranith Bottu (pbottu2)

Tony Tumba (ttumba2)

Davy Ji (davyji2)

Ron Basak (basak4)

CS410 Final Project Progress Report

Please upload your progress report to the Github repo shared on CMT. The progress report should give us an idea of how you're implementing your proposal. It should answer 3 main questions:

1) Which tasks have been completed?

- ❖ We researched common sentiment analysis techniques and libraries that are commonly used. Specifically, we found two libraries for our use: textBlob and vaderSentiment.
 - To determine which library to use, we first created two text files, each consisting of 30 sentences, having sentences that have positive and negative sentiment. These were generated ourselves since we know for sure if our analysis is correct.
 - As part of testing these libraries along with the text files, experimenting with the parameters and threshold was necessary to improve our results. In conclusion, using textBlob gave us 60 - 70% accuracy while using vaderSentiment gave us approximately 85% accuracy as shown in the source code, sentiment_experiment.py
- ❖ Researching various types of social media API, mainly twitter, facebook, instagram.
 - We have decided to use Twitter's social media API. Twitter was chosen due to its short tweets that are easier to analyse due to its character limit. Compared to Facebook posts or youtube comments, Twitter seems like an optimal social media choice. As for API usage, Twitter also is undergoing an API change, with its API v2, a lot of the bulkier parts of v1 have been updated and will be much easier to use.

2) Which tasks are pending?

- ❖ We still need to research the problems with social media scraping and any API constraints that may need to be taken into account when doing sentimental analysis. This will be one of the next steps we take in our research/planning process.
- ❖ Since we have recently decided to use Twitter's API, we would need to make an account via the generated token keys. Once we have this, we could further experiment with the API functions and limitations.
- ❖ Also, we will need to research what APIs and other polling software we will use to scrape public social media text. Finding tools that will be able to make our life easier will be essential, so making sure we find beneficial APIS/software will be important. This step as well as the designing process will determine how smooth the implementation will go.
- ❖ Finally, we still need to decide on how to design various functions to receive text data. This will be one of our most vital steps since how well we plan this will decide how smoothly the process of implementing it will go. We have begun this process already, but still need to finish it up and make sure it is done well.

3) Are you facing any challenges?

- ❖ As with any project utilizing other people's data, there are some privacy concerns regarding how that data is utilized. While there are Terms & Conditions that users agree to with signing up, we're not sure how far that covers our project's goal. As a result, the first challenge we faced (but managed to solve now) was ensuring the legality of how we're using the data.
- ❖ One challenge that we have is improving the accuracy of our sentiment analysis and robust testing of our model. Though we have a relatively high accuracy of 85%, we should certainly increase the number of sentences in our textfiles to reinforce our accuracy. Additionally, it is possible that the parameters and threshold we have chosen are not ideal. In order to improve our accuracy, we could keep experimenting or have a deeper understanding of the statistics and implementation behind the scenes.
- ❖ The second challenge we're working on right now is which API is the most beneficial for polling/scraping social media text. We've narrowed down the list to a couple that seem

good but have yet to finalize on one. Our main goal is to pick the one that makes our life the easiest and provides the most functionality. We'll likely have a solution for this soon.

- ❖ The last challenge we're working through is designing the structure of the various functions to receive text data. Our group is currently split on which algorithm would be best for executing our goal. As a result, the overall workflow is not yet finalized. We've started coding out small parts to test if our ideas work and plan to wrap this up later this week.