Team: ASA
Captain: Angela Luo (NetID: angelal6)
Sahithi Muthyala (NetID: sahithi4)
Ashley Tin (NetID: atin3)
Fall 2022
Professor ChengXiang Zhai

CS 410 (Text Information Systems) - Project Written Documentation

Overview of the function of the code (what it does and what it can be used for)

The code does a Twitter sentiment analysis on companies. This can be used for a number of use cases involving gaining an understanding of the public's perception of a company, including:

- From an individual standpoint:
    - Understanding company culture as a job applicant, interview candidate, or candidate with an offer from said company
    - Gauging perception of a company, the work that they are doing, and other relevant information when considering investing in a public company's stock
- From a company standpoint:
    - Gaining an understanding of what people think of one's own company to inform whether actions need to be taken to improve the state of external company perception
    - Learning about how the public perceives different competitors when doing a competitive analysis and aiming to learn more about the specific market that they are in

Examples of searches that may be helpful for users include "Company layoffs", "Company diversity", "Company stock", and "Company culture", where "Company" represents the name of a company. All of these searches are relevant for anyone who is looking to learn about a company. Our inspiration for building out this sentiment analysis specifically was helping Computer Science students quickly learn more about companies that they are seriously considering working for. Finding what a candidate considers to be a red flag through a quick search and browsing through results can save hours of their time and energy, as they can narrow their list of companies to focus on investing their time in while recruiting. We wish that something like this existed a few months ago when we were recruiting for full-time job positions.

Documentation of how the software is implemented

The software is implemented by:

- Pulling tweets via connecting to and leveraging the Twitter API. In order to do this, we made a Twitter developer account and got API keys.
- Retrieving tweets related to a keyword topic after making a request about a certain company.
- Using a Naive Bayes classifier and 14,000 sample tweets from the Natural Language Toolkit (NLTK) to train our sentiment analysis. Since we do not have strong personal computers available to train our sentiment analysis, we used Naive Bayes closely following this tutorial: https://www.digitalocean.com/community/tutorials/how-to-perform-sentiment-analysis-in-python-3-using-the-natural-language-toolkit-nltk.
- Leveraging the sentiment analysis to determine if searches regarding a company are positive or negative.
- Creating a word cloud of the tweets associated with the search query that the user enters. Users can use this word cloud to quickly determine the reliability of our sentiment analysis themselves instead of simply taking our word for it. We recommend that users use our sentiment analysis to gain a rough idea, but not as their only measure of if a company is positive or negative, as there is a lot more complexity to this.

A more comprehensive breakdown of our code will be provided in our video submission.

Documentation of the usage of the software (including documentation of usages of APIs and detailed instructions on how to install and run a software)

We used the Twitter API directly instead of Tweepy, which offered pre-build methods for the requests. Since Twitter switched its API format in November 2021, there is a lack of reliable documentation for Tweepy if we do not have a higher tier API access online. Users looking to extend upon our sentiment analysis should take into account their level of access for their specific use cases. The only action that they will have to take is obtaining a Twitter API authentication token, which requires them to create a Twitter developer account. More detailed instructions on how to do this can be found here in the Twitter developer docs: https://developer.twitter.com/en/docs/authentication/oauth-1-0a/obtaining-user-access-tokens. After they obtain this, they will need to add this token to our code in code block 2 on the line that says header = “” (we have left a comment on that line for users). In addition, users will have to install NLTK, the instructions for which can be found here: https://www.nltk.org/install.html.

Other than the imports and installations outlined in our Jupyter Notebook and above, users will not need to make additional installations in order to run our sentiment analysis. They will simply have to run the Jupyter notebook that we have provided and enter their search query in the line that we have specified for them to do so. This line is "query = " in code block 2.

Contribution of each team member

Each team member contributed equally to the project. The following is a breakdown of the split of workload:

- Angela
    - Code implementation of retrieving tweets and parsing them
    - Word cloud implementation
- Ashley
    - Code implementation of sentiment analysis
    - Integrating sentiment analysis code with tweet retrieval and parsing code
- Sahi
    - Finding resources for code implementation and pair programming with Angela
    - Writing positive and negative sentiment tweets for the dataset
    - All project written submissions and documentation