CS410 Course Project Proposal: Reddit WSB Stock Sentiment

Team Name: Blue Team

1) What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

   This will be an individual project by Kevin Cen, netID: kcen2.

2) What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

   I changed my topic due to challenges with trying to implement BM25 using a Chrome browser. I spent several hours trying to learn how to code in JS for a chrome extension and trying to index webpage so I could implement BM25 search. However due to time constraints, I recognized I was running out of time and needed to try something else. Fortunately due to the course project extension, I was able to switch.

   I've instead created a scraper that goes through comments/posts in the subreddit "r/wallstreetbets" based-off a ticker list provided by the user and scores the sentiment in comments that mention the ticker. This is an important problem because stock trading is increasingly influenced by the retail crowd and being able to scrape popular retail trading platforms for sentiment analysis might be useful in predicting future moves or finding topical tickers/stocks.

3) Briefly describe any datasets, algorithms or techniques you plan to use

   I use the VADER open-sourced tool: "VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is *specifically attuned to sentiments expressed in social media*."

   GIthub link: https://github.com/cjhutto/vaderSentiment

4) How will you demonstrate that your approach will work as expected?

   This will have to be a potentially future extension to try and backtest and see if sentiment scraped from reddit using VADER helps to predict stock prices.

5) Which programming language do you plan to use?

   I used Python.

6) Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

   This course project ended up taking a lot of time mainly because I tried unsuccessfully to implement BM25 using Javascript in a Chrome extension. I have had no experience with extensions, HTML or Javascript prior to this class and due to time constraints, I had to switch topics.