Project coordinator: Sita Jothishankar (slj2@illinois.edu), Asha Agrawal (ashaa2@illinois.edu)

What is your free topic?

Twitter sentiment analysis for stock price direction.

Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?

What is the function of the tool and its output?

This tool will scrape Twitter for a stock ticker name and then perform sentiment analysis to see how much positive and negative news there is regarding the term.

Who will benefit from such a tool?

Our main audience will be banks, investors, and shareholders. After finding the ratio of positive/negative news, they can incorporate this data into their views to make more robust and preemptive decisions for different commodities, stocks, companies, etc. This can help investors either buy low - sell high, or mitigate losses early on.

Does this kind of tool already exist? If similar tools exist, how is your tool different from them? Would people care about the difference?

We were not able to find an actual tool that found the sentiments, but found a study that showed that there is a correlation between between tweets and stock performance: https://towardsdatascience.com/stock-prediction-using-twitter-e432b35e14bd and http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf. With this supporting study, we believe that our project has a potential to make an impact.

What existing resources can you use?

https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vaderSentiment.py and https://github.com/satishrath185/Movie-Review-Sentiment-Analysis/blob/master/Sentiment%20Analysis.ipynb can give us a launching pad for a sentiment analyzer. We will be adding to it to create a more robust analyzer.

What techniques/algorithms will you use to develop the tool? (It's fine if you just mention some vague idea.)

We will use cross validation to create a sentiment analyzer on a dataset from kaggle: https://www.kaggle.com/yash612/stockmarket-sentiment-dataset

Once we have created the analyzer we will use the twitter API to perform sentiment analysis on current tweets to provide users with the most up to date information. As mentioned above, we will be using existing resources and studies to help us figure out which algorithms would be best to use.

How will you demonstrate the usefulness of your tool?

We will want to show correlation between stock price and the sentiment of news stories. With this correlation, we will be able to demonstrate the usefulness as the analyzer will provide our audience a preemptive measure for their stocks.

Which programming language do you plan to use?

Python

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.

There are 2 students on this team, so we will work on this project for 40 hours. We first have to research what algorithm we plan to use to create the analyzer and then we have to implement the algorithm. Once this is done, we have to test and train our analyzer. After this, we will use our analyzer to perform sentiment analysis on current tweets and create a user interface for the user to interact with the analyzer for most up to date information.

Each part of the project (listed below in the rough timeline) will be approximately 10 hours.

A very rough timeline to show when you expect to finish what. (The timeline doesn't have to be accurate.)

Part 1: Oct 26 - Nov 2	Research algorithms and learn more about sentiment analysis
Part 2: Nov 2 - Nov 16	Create and train sentiment analyzer
Part 3: Nov 16 - Nov 25	Perform sentiment analysis on current tweets using our sentiment analyzer and twitter API
Part 4: Nov 25 - Dec 9	Create user interface for user to interact with analyzer + write out documentation of our project