Project Title: To stock or not to stock.
Patrick Du, Kiseok Yang, Chuanyao Lin

Project Description:
The goal of this project is to predict the specific price of a stock at any given point based on previous existing data. We plan to categorize stocks into 3 groups, stable stocks, new stocks and volatile stocks. First and foremost we will be scrapping news headlines off Bloomberg with Bloomberg's API via rapidapi. Next we will formulate Sentiment Analysis values on the headlines as well as creating bigram/trigram sentiment dictionaries. These dictionaries will be analyzed by a Sentiment Analysis algorithm to determine whether a given headline for a company is doing good or bad. As soon as the Sentiment Analysis finishes making a prediction, we will add the values into the stock as a feature. We will then use Feature Selection via SVM to improve the accuracy of our prediction. Finally we will apply LTSM on the whole dataset to find whether this stock should be invested in. The scope of the project are as follows

- Testing different ways to extract dynamic data in realtime to formulate a prediction.
- Testing different Sentiment values to improve Sentiment dictionaries reliability.
- Testing Feature Selection to improve the accuracy of our prediction.
- Reliably predicting the price of a stock with LTSM.
- Potentially improving the algorithm by implementing ways to adjust for real world events.
- Potentially using global data to make predictions while the US market is considered closed.

Each team member's role
As of right now the roles are:
Patrick Du: Coordinator and working on Sentiment Analysis
Kiseok Yang: Working on LTSM
Chuanyao Lin: Working on Feature selection with SVM
Everyone:

- Organize the dataset to be usable in the code
- Manually give sentiment values to headlines and bigram/trigrams.
- Build code
    - -implement algorithm
    - -Result error check
    - -Visualize the result
- Report
- Presentation

Dataset: The fundamental data is from kaggle.com, which contains Dates, Open, High, Low, Close, adj Close and Volumes and update the recent fundamental data and other useful data is from financialmodelingprep.com. The dataset is taken from stocks of three different categories, which are stable, new and volatile. Stable stocks are those that see plateaued activity (very little dips and hills), new stocks are those who have been in the market for under 3 months and volatile stocks are those that have huge up and down spikes.

Papers:

We will be using the following papers as a guide, reference and challenge to create and improve on several algorithms.

- This one is used for SVM and Feature Selection:
  Lin, Yuling, et al. *An SVM-Based Approach for Stock Market Trend Prediction - IEEE Conference Publication*, 2013, ieeexplore.ieee.org/document/6706743.
- This one will be used for LTSM:
  Nelson, David M. Q., et al. "(PDF) Stock Market's Price Movement Prediction with LSTM Neural Networks." *ResearchGate*, www.researchgate.net/publication/318329563_Stock_market's_price_movement_prediction_with_LSTM_neural_networks.
- This one will be used for globalizing all financial markets :
  Shen, Shunrong, et al. *Stock Market Forecasting Using Machine Learning Algorithms*. cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf.
- This one will be used for Sentiment Analysis:
  Kalyanaraman, Vaanchitha, et al. "Sentiment Analysis on News Articles for Stocks." *Sentiment Analysis on News Articles for Stocks - IEEE Conference Publication*, 2014, ieeexplore.ieee.org/document/7079267

Timeline for the project

- Scrapping Bloomberg (Done by 11/4)
- Sentiment Analysis (Done by 11/10)
- LTSM (Done by 11/8)
- Feature Selection (Done by 11/18)
- Visualization (Done by 11/25)
- Preparing for final report and presentation (Done by due date)

Describe our demo and other deliverables

- The demo will contain a video of us running both algorithms as well as showing the results.
- The final presentation will explain all of the details in our study, show our results and future direction from the results.

Code

- There will be 3 segments of code. One segment will focus on creating objects specific to the stock as well as the data we need. Another segment is focused on the SVM approach. The final segment will be focused on the Neural Network approach. This can change.