

CS230 Project Proposal: Stock Prediction using Deep Learning

Category: Natural Language Processing

The problem to be investigated is stock price and volatility prediction by extrapolating data obtained from relevant sources such as sentiment from financial news headlines, social media blogs, previous stock history, and then integrating the data as input to a deep learning model to predict the outcome of any stock ticker symbol in upcoming days. It is an interesting problem because stock prices are generally difficult to determine, but using automated prediction can be beneficial for investors to have an idea of upcoming trends in stocks before the price change actually occurs.

The main challenge is obtaining a substantial amount of data from recent sources in order to build an accurate training model. Also, the highly-parameterized nature of influences on the stock market pose a great challenge to accurately predict a stock price change at any given time.

The datasets I plan on using are from AlphaVantage which provides free APIs for stock data, as well as archived articles from Reuters and Bloomberg based on data previously collected from Ding et al. (2014). Some examples to be used for training and test data may include recent sources obtained from personal web scraping tools.

Since the task mainly relies on evaluating text and is therefore a natural language processing problem, word embeddings from Word2Vec or GLoVe combined with an RNN or CNN architecture with long-term memory capability will most likely be used. Some existing implementations I found mainly train on article headlines, but I also plan to cross-reference data from relevant social media posts and possibly numerical data from past history of particular stocks obtained from AlphaVantage in order to gather data from other potential influences.

“Deep Learning for Event-Driven Stock Prediction” by Ding et al. (2015) may provide useful references for context of and datasets and architectures used.

Evaluating the results qualitatively can be plotted comparing the accuracy of stock price over a number of days vs the model’s prediction. Quantitatively, the results can be evaluated using the Standard Measure of Accuracy (Acc) and Matthews Correlation Coefficient (MCC) on the S&P 500 index prediction and stock prediction.