Eric Escareno
Christian Galleisky
Irvin Budwal
Joseph Lopez

CS478 Mid-Progress Report: Stocks Deep Learning

Our team has decided that we will try to model the emotional state of the Nasdaq stock market in order to understand when the market is emotionally overbought or oversold. What we aim to do is to understand when we should expect some up or down movement in a particular security(equity/stock), so as to not be shocked when it does actually occur. At this point these are tentative goals; We have never done any work in Machine Learning and thus, have almost no idea of how big any one idea is. We will do our best to deliver a quality product/analysis by semester's end. Please bear with us and grade us on progress rather than perfection.

Market psychology paired with deep learning can give us greater insight into how and when we should make trades. This is interesting to our group as most of us have very little investing experience so if novices can figure out how to leverage deep learning effectively there are many more possibilities. Readings that our team will utilize to gain contextual understanding in order to tackle this project are: *Getting Started in Technical Analysis* by Jack Schwager, and *Technical Analysis* explained by Martin Pring. Both books give detail on technical analysis which can be used to make better predictions in stock price. A large portion of the data and formatting we will use can be found at:

https://www.kaggle.com/jacksoncrow/download-nasdaq-historical-data/notebook

We will try to get current data by either web scraping on stock websites where possible or trying a script that collects the data.

https://www.kaggle.com/datasets/jacksoncrow/stock-market-dataset/discussion/188637

If we cannot find a means to quickly collect new data then we will use existing data to build upon. We will use the Monte Carlo simulation method. There are existing implementations of this method which we can build off of and tweak. Most likely this will require adding different forms of data to include as context. We will evaluate our results by checking whether our model makes accurate predictions. Most likely, we will be utilizing cross validation as a performance metric. This will allow us to test our model under certain stressors, so that we can determine prediction accuracies. Essentially, by dividing data from our chosen subset into two segments, one segment is then used to train our model, while we use the other to compare it to. (Comparing independent vs dependent variable). We will need to learn technical analysis, which market psychology is an aspect of, to make our own predictions to compare against. Our results should be representable as plots and we should be able to compare the accuracy of our predictions across sample sizes to tweak our activation function. Ultimately, the goal is that our market analysis model will help streamline investment and trading for newcomers and people unfamiliar with that environment, for users to maintain their focus on spreads and trends. There are existing implementations which we can build off of and tweak. Most likely this will require adding different forms of data to include as context. These can include:

Classical Economic indicators; capacity utilization, industrial production, inventories growth/contraction, unemployment rate, participation rate, savings rate, income growth rate, disposable income growth/contraction, etc…

Classical Business metrics: Price to earnings, price to earnings and growth, gross margin, operating income, Earnings before taxes interest and deductions (EBITDA), profit margin, Earnings per share, long term debt, assets, liabilities, assets minus liabilities, etc…

Classical stock chart ideas; shapes of stock price graphs and shapes of various stock chart metrics like RSI and Stochastic Oscillator, etc…

We will evaluate our results by checking whether our model makes accurate predictions. We will need to learn technical analysis, which market psychology is an aspect of, to make our own predictions to compare against. Our results should be representable as plots and we should be able to compare the accuracy of our predictions across sample sizes to tweak our activation function. The ultimate goal would be to make the model as accurate as possible.

Essentially, we have made some changes to our project and its direction that we believe will be beneficial to our timeline. We have focused on reducing the approach of “emotional state” of the stock market as a means to replicate neural networking. On top of that, deciding to change our approach from exclusively following current data, to older backlogs containing lengthier existing data so we can better make analysis.

Deciding between both datasets was a difficult choice. Historical daily prices and volumes of all U.S. stocks and ETFs from (Kaggle) included a data set that only went up to 2017. It was very in depth and expansive, however the data included was static, brute forced to maintain numbers up to the year 2017. Historical daily prices of Nasdaq-traded stocks and ETFs also from (Kaggle) was the data set that we decided on. This is mostly because the dataset is extremely granular, the last dynamic update was in 2020, the contributors to the project included through and extensive documentation, and included was a data collection script that essentially allows us to update the stock data to more current “quarterly” fixtures.

The basis for our model will essentially include close training and prediction according to moving averages. (This data can be cross referenced from data analyzed in the 1980's to 1990's)

The main force of our focus will be on developing a machine that will utilize LSTM. Long short-term memory will allow us to predict stock trends from specific sequences of variable lengths. From related works that have attempted this, we are confident in generating an algorithm that can successfully and effectively predict the movement of the stock market. This model would then be adjusted to individual stocks and stock exchanges, depending on the user's desire. The idea is that it can be applied to any stock and be an effective tool in future predictions. (Using it to predict apple stock, microsoft, etc.)

A latest iteration that we have seen utilized an LTSM algorithm to closely predict the trends from the IBM stock, with very minimal deviations. It's not 100% accurate, but the shape prediction is very close. We believe this will be an extremely useful tool for users that are in need of assistance for making smart buying choices, to help with involvement in the stock market.

Our Group plan for final deliverables:

- Abstract: Irvin
- Introduction: Chris, Eric, Joe, Irvin
- Related works research: Eric
- Data and processing: Chris, Eric, Joe
- Methods/approach: Eric, Irvin
- Experiments: Chris
- Conclusion: Joe