

Proposal: [Maximizing ROI using Machine Learning]

DATA 450 Capstone

[Matthew Zimmermann]

February 8, 2024

1 Introduction

Since the beginning of the financial markets, people have been trying to find every possible way to maximize a return on their investments while maintaining a level of risk management that allows them to trade safely. The golden question of, “can the stock market be predicted?” has been asked and studied for a long time. It seemed as if predicting the stock market was an impossible feat, and can still very much be so, however, with the introduction of machine learning into the financial world, it seems that the answer is getting closer and closer to being solved. The introduction of machine learning is revolutionizing the financial world, and is allowing investors to make more informed investment decisions in a space that is extremely complex. In this analysis, I will utilize an LSTM model to try and predict future price action of a stock like Apple(AAPL) to the best of my ability, to see if even a small amount of machine learning can increase the ROI of an investment by a considerable amount.

2 Dataset

The dataset I will be using was found on NASDAQ’s official website. This dataset is not available for download, instead it can be obtained through the NASDAQ python package. From there I can pull specific data that I need to create my own dataset. The data is free in the package, however in order to use it I must obtain an API key from their website by creating an account. The variables I plan to use in my analysis are the following: * date * open_price: price that the stock opened at on the given day * close_price: price that the stock closed at on the given day * high_price: the highest price that stock reached on the given day * low_price: the lowest price that stock reached on the given day * volume: number of shares traded on the given day

“AAPL.” Nasdaq, www.nasdaq.com/market-activity/stocks/aapl. Accessed 8 Feb. 2024.

3 Data Acquisition and Processing

I will obtain this data through the NASDAQ python package. The data is not downloaded off of their website, instead I will be able to pull whatever data I need from the python package. From there I will need to delete any unwanted variables, replace any missing values, if any, and change variable names. There could be more data processing that I am unsure of, however it should be simple and straightforward data processing due to them keeping their data well maintained and up to date. The package allows users to gather all variables according to the given ticker symbol in the form of a dataset.

4 Research Questions and Methodology

1. How well can we predict the future price action of AAPL? To answer this question, after gathering the data necessary, I will create an interactive candlestick graph that allows the reader to explore the price of AAPL on any given day. I may create moving averages and graph those aswell. Then, the creation of training and testing sets in preperation for the LSTM model will need to be made. After this, normalizing the data and training the model. After the model is trained I will explore the results and graph it visually for the reader as well.
2. How much better ROI is attainable using the trained model, than if not? To answer this model I will provide the reader with two seperate trading strategies. The first would be an average long term hold on AAPL spanning the course of a year. The initial purchase happening on Jan 1. and the position closed on Dec. 31. The second strategy will be using the model I trained and tested to make trades based on the results. (For example, if the model predicts the price a week from the initial purchase to be lower than the purchase price, the postion will be closed, and if the predicted price is higher than the purchase price from the previous purchase the postion will remain open.) I will then make a side-by-side bar plot showing the returns of both strategies both taking into consideration taxed and non-taxed, since long term and short term trading are taxed differently.

5 Work plan

Week 4 (2/12 - 2/18):

- Data gathering, figuring out NASDAQ python package (4 hours)
- Data tidying/processing (2 hours).

Week 5 (2/19 - 2/25): * Creating an interactive candlestick graph of APPL stock (4 hours)
* Moving averages and add them to separate or interactive graph (3 hours)

Week 6 (2/26 - 3/3): * Create training and testing sets (1 hour) * Normalize data and train the model (2 hours) * Graph the model results along with actual price, interpret results (4-5 hours)

Week 7 (3/4 - 3/10):

- Presentation prep and practice (4 hours)

Week 8 (3/11 - 3/17): *Presentations given on Wed-Thu 3/13-3/14. Poster Draft due Friday 3/15 (optional extension till 3/17).*

- Poster prep (4 hours)
- Presentation peer review (1.5 hours)

Week 9 (3/25 - 3/31): *Final Poster due Sunday 3/31.*

- Peer feedback (3.5 hours)
- Poster revisions (3.5 hours)
- [Do not schedule any other tasks for this week.]

Week 10 (4/1 - 4/7): * Start question 2 * Create necessary charts, yearly return without model * Try to get as far as possible figuring out how to gather the second trading strategy I mentioned previously (together 7 hours)

Week 11 (4/8 - 4/14): * Finish question 2 * plot returns of trading strategy 1+2 side-by-side, including and excluding tax (together 7 hours)

Week 12 (4/15 - 4/21): * Write up and interpret results of entire analysis * Make graphs and charts better, easier to understand * Any finishing touches (together 7 hours)

Week 13 (4/22 - 4/28): *Blog post draft 1 due Sunday night 4/28.* [All project work should be done by the end of this week. The remaining time will be used for writing up and presenting your results.]

- Draft blog post (4 hours).

Week 14 (4/29 - 5/5):

- Peer feedback (3 hours)
- Blog post revisions (4 hours)
- [Do not schedule any other tasks for this week.]

Week 15 (5/6 - 5/12): *Final blog post due Weds 5/8. Blog post read-throughs during final exam slot, Thursday May 9th, 8:00-11:20am.*

- Blog post revisions (2 hours)
- Peer feedback (2 hours)
- [Do not schedule any other tasks for this week.]

6 References

[The bibliography will automatically get generated. Any sources you cite in the document will be included. Other entries in the `.bib` file will not be included.]