Course Project Proposal

CSCI-4502

William Ruiz
Computer Science
University of Colorado Boulder
Boulder Colorado USA
wiru2611@colorado.edu

Kasper Seglem
Computer Science
University of Colorado Boulder
Boulder Colorado USA
kase1828@colorado.edu

Angelo Vacca
Computer Science
University of Colorado Boulder
Boulder Colorado USA
anva5757@colorado.edu

INTRODUCTION

For our project, we will be analyzing the stock history of the current S&P 500 companies. We hope to find patterns and determine correlation between stocks, using this information to make predictions. As an example, stocks in the same industry may often go in the same direction. This data would be similar to itemsets, where the stocks with similar movement would be in the same set.

KEYWORDS

Finance, stocks, S&P 500, predictions, trading, algorithms.

RELATED WORK

1 Hudson River Trading [1]

Hudson River Trading is a firm located in New York that specializes in algorithmic trading of stocks through rigorous analyses and attempting to predict future price movements based on live data. This firm utilizes their trading algorithms to invest your money in order to beat the S&P 500 average.

2 J.A.R.V.I.S [2]

JARVIS is similar to Hudson River Trading, however the software utilized runs on top of the TradingView platform to give the user a more hands-on experience versus a company doing the trading for you.

We will try to take inspiration from both of these algorithms, while also developing our own methods. However, we hope to mainly rely on our own data findings.

PROPOSED WORK

1 The Dataset

We will be using a dataset containing the historical data for each of the current S&P 500 companies. It includes six different values, all of which may be used for our algorithm. The dataset is publicly available at:

https://www.kaggle.com/datasets/camnugent/sandp500?resource=download.[3]

Kaggle is a subsidiary of Google, where people can post datasets. It was made primarily for data scientists and machine learning practitioners.

It contains data from the past five years. It is in comma separated value format and contains the following columns for each day of the market:

- Date - in format: yy-mm-dd
- Open - price of the stock at market open (this is NYSE data so all in USD)
- High - Highest price reached in the day
- Low Close - Lowest price reached in the day
- Volume - Number of shares traded
- Name - the stock's ticker name

We will likely start with the most recent activity and work our way backwards.

2 **Correlation between stocks**

One idea we have is to study the correlation between individual stocks. If a strong correlation is proven, we can expect the stocks to perform similarly and use that information to make predictions. This is something that can be seen with stocks in the same industry. It is often the case that these stocks move together, or at least have some shared qualities.

3. **Predictive Algorithm**

The main goal for the project is to create a predictive model for the stock market. This could include using a stocks own historical data or using the data of its closely correlated stocks. We may also include elements from other algorithms if we decide they may be useful.

EVALUATION

We can evaluate our work by analyzing our predictions and whether they correlate/match stock changes or patterns for that day. We can also compare our model with other models such as Hudson River Trading or J.A.R.V.I.S and see how our model compares. While we do not expect to have accuracy close to matching that of these professional made models, they can give us a good baseline as to what a well made model is able to do and predict. Another way we can evaluate our work is by comparing our predictions to randomness. This can give us insight into knowing whether our model can actually predict trends and changes in the stock market, and is not simply a random algorithm.

Our metrics for this evaluation will be the accuracy of our predictions. More specifically, it will be the percentage change predicted by the algorithm, compared to the percentage change of the actual stock being tested. Percent error ranges will also be used when evaluating our model, and statistical tools like standard deviation can tell us how off our predictions are. Comparing our error ranges to that of randomness and other models will also allow us to evaluate how much better or worse our predictions are compared to other approaches and algorithms.

MILESTONES

1 **Week 7**

Collect past and current stock data from the S&P 500. This includes finding an up to date dataset with the history of each stock in the S&P.

2 **Week 10**

Study developed trading algorithms and models used to help predict stock movements. This includes learning more about the Hudson River Trading algorithm and the J.A.R.V.I.S. algorithm.

3 **Week 12**

Project checkpoint report. Put together a report to document current progress.

4 **Week 13**

Study and mine data in order to gain a better understanding of historical stock data by finding trends, patterns and outliers in movements. Use the S&P 500 data and begin to mine patterns and correlations.

5 **Week 15**

Use data from the previous milestone to create a rudimentary predictive model/algorithm based on patterns and trends from mined data that allows fictitious stock trading to happen.

6 **Week 16**

Testing, final project and final report. Study the accuracy of our new algorithm and document it in the final report.

REFERENCES

[1] Hudson River Trading - HRT. n.d. *Home » Hudson River Trading - HRT*. [online] Available at: <https://www.hudsonrivertrading.com/> [Accessed 6 October 2022].

[2] Jarvis-algo.com. n.d. *J.A.R.V.I.S - trading algorithm*. [online] Available at: <http://www.jarvis-algo.com/> [Accessed 6 October 2022].

[3] Nugent, C., 2018. *S&P 500 stock data*. [online] Kaggle.com. Available at: <https://www.kaggle.com/datasets/camnugent/sandp500?resource=download> [Accessed 6 October 2022].