

Group 84 Project Proposal

Jordan Coleman, Aayush Dixit, Prince Fodeke, Elias Izmirlian, Gabriel Montes

Background

Machine Learning applied to the stock market is by no means a new area due to the profitability of models that can predict stock prices. Various algorithms have been tested against each other in performance and computational efficiency. In Huang *et al.* (1) SVM, Linear Discriminant method, Elman Backpropagation Neural Networks and Quadratic Discriminant were compared in their ability to predict financial trends by evaluating weekly trends of the NIKKEI 225 index. The results found that SVM was the most effective and computationally efficient method. In Nabipour *et al.* (2) multiple algorithms were employed to tackle a similar problem, and the use of a binary representation for the data was compared with a continuous representation. The experiment found that the ensemble achieved up to 67% accuracy in predicting stock trends, whereas with binary data it improved to 83% accuracy. Lastly, Hegazy *et al.* (3) used Least Squares SVM algorithms integrated with particle swarm optimization to accurately predict stock market prices. This model was compared with a neural network and was found to have better prediction accuracy through the use of multiple financial datasets. While this field has been explored, we hope to gain a better insight on how using machine learning to predict stock prices could directly affect companies in the S&P 500.

The dataset we plan to use contains data on the companies in the S&P 500 over the past few decades. Specifically, some of the features that we plan to use include opening price, closing price, lows, highs, and date. We aim to use this data in order to train a model to predict weekly S&P indices.

Problem Definition

There are many different machine learning and deep learning methods used to predict stocks. We will test several of these methods to determine which will be best in predicting the day to day closing prices of the S&P 500 index, which is regarded by many as the best valuation of the American Stock Market.

Methods

The methods we will use to predict the values of the S&P 500 index: Support Vector Machine, Elman Backpropagation Neural Networks, and Long Short Term Memory Network, each will use historical data to train the model using the features mentioned above. The data will be taken from Date to Date and was collected from Datahub.io. The data will follow an 80-20 training-testing split, and we will use the root means square error (RSME) metric to evaluate the accuracy of each method.

Potential Results/Discussion

Potential results will attempt to provide predictive insights into future movements of these major index funds within a non-insignificant relationship to these other indicators. This could also be used to verify commonly accepted correspondences between US economic indicators such as gold prices and the S&P 500. These results could be used to further solidify or challenge the common assumptions between common indicators and the variety of ML techniques could allow us to assess the accuracy and precision of the different methods used.

Contribution Table-Project Proposal

Team Member	Contribution
Elias Izmirlian	Wrote introduction and researched past uses of machine learning in stock prediction and provided possible algorithms to be used.
Aayush Dixit	Researched background further, namely further research into specific algorithms in training past models as well as some background on our possible dataset.
Jordan Coleman	Provided a formal problem definition that outlines the main goal of the project as well as further outlined the methods to be used.
Gabe Montez	Discussed potential results and possible impacts.
Prince Fodeke	Created GitHub page, set deadlines on gantt chart and created contribution table.

References

1. W. Huang, Y. Nakamori and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine", *Comput. Oper. Res.*, vol. 32, no. 10, pp. 2513-2522, Oct. 2005.
2. Nabipour, M., Nayyeri, P., Jabani, H., S., S., & Mosavi, A. (2020). Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. *IEEE Access*, 8, 150199–150212. <https://doi.org/10.1109/access.2020.3015966>
3. Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. *arXiv preprint arXiv:1402.7351*.