HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

Stock Price Predictions with A Linear Model and Neural Network

Ivan Ma[1] and Odysseas Drosis[#]

[1]Irvington High School, USA
[#]Advisor

ABSTRACT

The stock market is important for the economy: not only does it provide ways for businesses to raise capital, but it also allows individuals to build wealth. This paper focuses on building an artificial intelligence (AI) model that can predict the prices of stocks based on historical prices. The data used to train the AI model was collected via Yahoo Finance, and the result was a model that could predict the prices of shares of different companies with a mean square error (MSE) of less than 100.

Introduction

The stock market is where investors buy and sell stakes in companies. It has many uses and has a great impact on the economy and individuals. Because the stock market is vital to the economy, it is critical to understand the various purposes of it and its implications on the overall economy.

One purpose of the stock market is for businesses to raise capital. When a company goes public by offering partial ownership to the public, it can acquire a lot of funding. This funding can be used for various purposes, such as expansion, research and development, and repayment of debt. Thus, the stock market acts as a tool for economic growth and job creation by enabling companies to invest in their operations.

Another purpose of the stock market is wealth creation. The stock market offers opportunities for investors to build wealth over time. When investors buy shares in a company, they become shareholders of the company and can benefit as the company grows. Investments in high-demand stocks can generate significant returns in the long run and help individuals achieve their financial goals. Additionally, investments contribute to price discovery. Keeping track of the buying and selling of stocks helps determine the supply and demand of shares of certain companies, which is one of the determining factors of fair pricing for individual shares of companies. Price discovery is also important because it is one factor that investors consider when purchasing stocks.

Apart from helping businesses and investors, the stock market is also important because it can be used to determine the state of the economy. Periods when stock prices are increasing all around (bull markets) are an indication of a healthy economy and economic expansion, while periods when stock prices are plummeting (bear markets) indicate challenges in the economy. Therefore, the stock market is an important tool for analysts to track the health of the economy.

With that being said, investing in the stock market comes with risks. Prices can suddenly go up or down for many reasons, including changes in interest rates, corporate earnings, supply and demand, and geopolitical events. The stock market's volatility can provide opportunities but also pose risks. A long-term investment perspective is often used to mitigate risks associated with stock market investments. This is the purpose of this study: to use historical stock price data to train an AI model to predict future stock prices.

Data Set, Input, and Output

The data used was stock price data from Yahoo Finance. Google Colaboratory was used as a tool to import the data and make it usable as training and testing data for the price prediction models. The data included the stock price (USD) for Apple, Amazon, Meta, Tesla, Microsoft, and Walmart for 1,258 days. The Python library Pandas was used for data cleaning and analysis, and NumPy was used to store the data in arrays. The data was sorted in two arrays for the models. The first row of the first array would contain the prices for days one, two, and three; the second row would contain the prices for days two, three, and four; and the third row would contain the prices for days three, four, and five, and this would continue until the second to last day. The second array contains just one row with "outputs" for the first array. Each entry in the array is the stock price the day after the last day in each row of the first array. For example, since the last price for the first row of the first array was the price for the third day, the first entry of the second array is the price for the fourth day. This continues until the last day. The data in the two arrays would then be sorted into training and testing data sets ready to be put into the models. In this paper, the testing set consisted of 33% of the data, and the rest went into the training set. The models will then produce a new list of predictions based on the test data set after training with the training data set. The model will also calculate and return an MSE based on the accuracy of each prediction made.

Models

One of the models used in this paper was a linear model, which works with small sets of data. A linear model uses linear regression to estimate the relationship between an independent variable and a dependent variable, both of which must be quantitative. On a graph, this estimated relationship would be the line of best fit on a scatter plot, and this line would have the lowest mean difference between the estimated and actual values. Higher-order models (quadratic, cubic, quartic, etc.) can produce more accurate estimates, but a linear model is enough to produce satisfactory results for predicting stock prices. Figure 1 is a visual example of linear regression using the data of the prices of Apple stocks collected from Yahoo Finance. Based on this graph, it seems that the price of individual shares of Apple tends to increase as time goes on.

The other model used in this paper was a neural network, which is often used for larger sets of data. A neural network consists of layers of interconnected neurons, or nodes, similar to the human brain. In a neural network, data moves in one direction between layers, and each neuron will be assigned a value. For the first layer, the associated values for each node are input values, and the last layer consists of the output. For the layers in between, which are called hidden layers, nodes are assigned a value based on the sum of the products of the values assigned to the nodes in the previous layer and the associated weight of the nodes. The weight is a scalar and controls each value's impact on the output. A larger weight means a larger impact, while a smaller weight means a smaller impact. Neural networks can have several layers, and although more layers usually lead to improved accuracy, an increase in layers will also add to the computational time. Figure 2 visually displays the connections between nodes in the different layers of a neural network. The lines between each node represent the value assigned to the previous node multiplied by an associated weight. In the end, there is a single output for the entire neural network. For the purpose of this paper, the output will be the prediction of the price of a stock on a certain day.

Using these two models, a stock trading bot was created. It uses a weighted average of the predictions of the linear model and neural network to determine a daily prediction, and based on whether or not the stock price is predicted to increase or decrease and how much the stock price is predicted to increase or decrease, the bot would sell a certain amount of stocks. If the stock price is predicted to increase the next day, then the bot will buy stocks based on the predicted percent increase of the stock price on the current day. If the stock price is predicted to decrease the next day, then the bot will sell the stocks on the current day. The initial budget was 5,000 USD, and when tested on trading Apple stocks, the trading bot managed to gain a profit of 3,724.12 USD. Although the criteria for buying and selling stocks was not the most optimal, the bot still managed to earn 74.48% of the original budget.

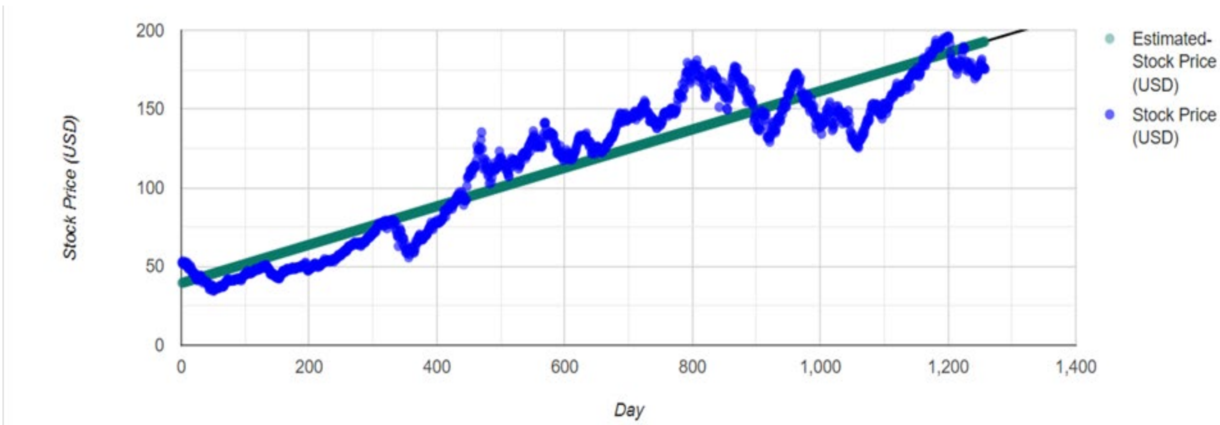

Figure 1. Stock Price Prediction Model for Apple for 1,258 Days

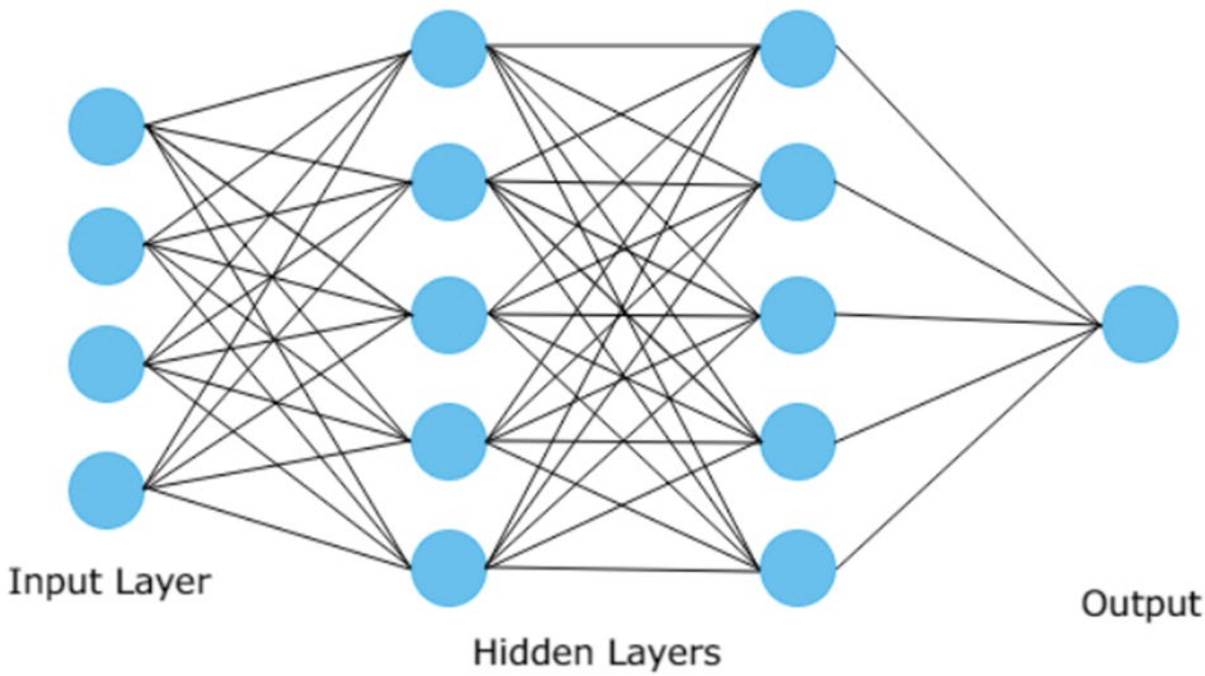

Figure 2. A visual representation of a neural network

Results

As shown in Table 1, it seems that the neural networks consistently perform slightly worse than the linear models. However, the difference is not large, and both models perform similarly for the same companies. They performed best when predicting the stock prices for Walmart and performed worse when predicting the stock prices for Tesla, both with the test set. Additionally, both the linear model and neural network exhibited subpar performance when predicting the stock prices for Tesla and Meta. This is due to the volatility of the prices for stocks of these two companies. In the period in which data was collected for the stock prices of Tesla and Meta, there were bull and bear markets, which caused the prices to deviate greatly from the predicted prices of the models and also greatly increased the MSE. It is especially clear when looking at the price points on a graph.

Table 1. A table of the MSE for the predictions of neural networks and linear models for Apple, Amazon, Meta, Tesla, Microsoft, and Walmart

Company	MSE with Linear Model (Test set)	MSE with Neural Network (Test set)
Apple	5.71	5.79
Amazon	9.24	10.00
Meta	33.86	34.92
Tesla	65.78	66.28
Microsoft	15.83	16.89
Walmart	2.66	3.45

In Figure 3, it is clear that Tesla stock prices started to become very volatile after day 500. After day 500, there were long periods where the actual price was well above or below the predicted price. On the other hand, the models performed very well when predicting the stock prices for Apple, Amazon, and Walmart. As opposed to the unstable trends of Tesla and Meta stock prices, the prices of Walmart, Amazon, and Apple stocks had very low volatility and followed a nearly straight line throughout the period in which data was collected for the stock prices. As seen in Figure 4, the price rarely deviates from the line of best fit by a large margin. From the results of the models, it is possible to conclude that linear models and neural networks work best when trends are mostly linear and relatively non-volatile. When a set of data has fluctuations like the stock prices for Tesla and Meta, it is better to use higher order regression so that the line of best fit and follow the trends closer and achieve a lower MSE.

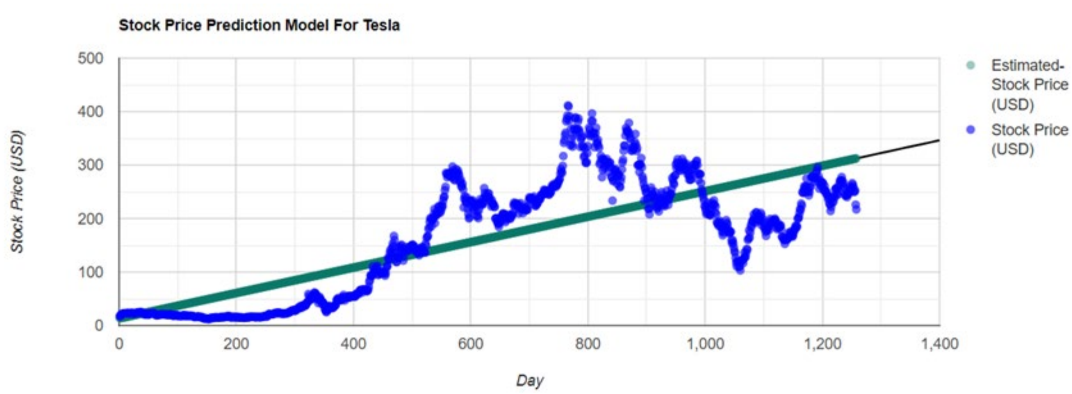

Figure 3. Stock Price Prediction Model for Tesla for 1,258 Days

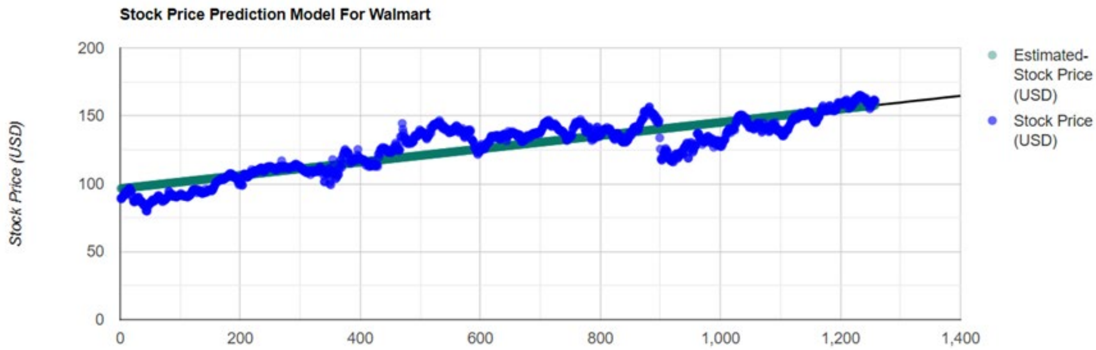

Figure 4. Stock Price Prediction Model for Walmart for 1,258 Days

Discussion

One of the main challenges in this paper was figuring out a way to make the linear model and neural network work well on more volatile stock prices such as those of Tesla. In such cases where there are sudden occurrences of volatility in stock prices, little can be done about the performance of the linear model since there will be a lot of error no matter what if it disrupts a stable trend. As for the neural networks, the volatility would need to occur in a predictable manner, and more data would be needed for the neural network to make accurate predictions even with fluctuating prices. However, it also might not be feasible to collect too much data since previous trends in prices may be vastly different from those in more recent years. It may be better to just use a higher order regression model when there is a smaller set of data.

Conclusion

This paper makes a linear model and neural network to attempt to predict the prices of stocks for six different companies. Unfortunately, both the linear model and neural network severely underperformed when trying to predict the stock price for Tesla and Meta, which both had very volatile stock prices. Unsurprisingly, though, both performed well when predicting prices for Walmart stocks, which had an almost-linear trend throughout the period in which the data was collected. While a higher order regression would work much better for more volatile prices, such a model would likely be unable to predict further unique trends in prices. Further research would be required to make a model work for predicting volatile prices and sudden trends with small pools of data, a model or algorithm that can quickly catch on to certain trends and adjust to those trends in order to make accurate predictions.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

1. Prasad, Akhilesh & Seetharaman, A.. (2021). Importance of Machine Learning in Making Investment Decision in Stock Market. Vikalpa: The Journal for Decision Makers. 46. 025609092110599. doi.org/10.1177/02560909211059992

2. Mokhtari, Sohrab & Yen, Kang & Liu, Jin. (2021). Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning. International Journal of Computer Applications. 183. 1-8. doi.org/10.5120/ijca2021921347
3. Awan, Mazhar & Rahim, Mohd & Nobanee, Haitham & Munawar, Ashna & Yasin, Awais & Zain, Azlan & Javed, Mazhar. (2021). Social Media and Stock Market Prediction: A Big Data Approach. Computers, Materials & Continua. 67. 2569–2583. doi.org/10.32604/cmc.2021.014253
4. Joshi, Srivatsa & Kumar, Vishwas & Venkataramanan, Vishaka & C S, Kaliprasad. (2023). A Review on Neural Networks and its Applications. Journal of Computer Technology & Applications. 14. 2023. doi.org/10.37591/jocta.v14i2.1062
5. Vadlamudi, Siddhartha, (2017), Stock Market Prediction using Machine Learning: A Systematic Literature Review, American Journal of Trade and Policy, 4, issue 3, p. 123-128. doi.org/10.18034/ajtp.v4i3.521
6. Mehar Vijh, Deeksha Chandola, Vinay Anand Tikkiwal, Arun Kumar, Stock Closing Price Prediction using Machine Learning Techniques, Procedia Computer Science, Volume 167, 2020, Pages 599-606, ISSN 1877-0509, doi.org/10.1016/j.procs.2020.03.326.