Using Machine Learning to Predict Future Stock Prices

Sriya Amineni

Pennsbury High School

ABSTRACT

The volatile nature of financial markets presents a formidable challenge in accurately forecasting equity returns and stock prices. Artificial intelligence (AI) and machine learning are increasingly being employed to predict future stock prices and reduce risks, enabling more profitable investment decisions. In this paper, we employed a few AI regression and classification models to tackle the challenge of predicting future stock prices These models utilized a historical time-series dataset of the opening and closing prices of stock over the last five years to generate predictions that were eventually compared with actual prices The models were evaluated based on their regression or classification score, measured as a percent of accuracy. Among the models tested the MLP classification and linear regression models were the most effective predictors.

Introduction/Background

Since the establishment of the first modern stock market in Amsterdam, Netherlands during the 1600s, the platform has become an integral part of the economy all over the world. Currently, there are 60 major stock exchanges around the world and 11 different sectors: Information Technology, Consumer Discretionary, Health Care, Materials, Real Estate, Industrials, Financials, Energy, Consumer Staples, Utilities, and Communication Services. A company belonging to any one of these sectors has the ability to go public. Any institution that goes public has the ability to issue stock (the unit of stock is referred to as shares), which gives a market participant (a buyer or a seller) a small fraction of the company's equity ownership. Through this, businesses are able to raise funds to use at their discretion, while shareholders - the investors - have the ability to participate in shares appreciation or through the company's payment of dividends. Being able to predict the future value of the stock prices can be incredibly profitable to all those who invested in itt. Thus, studying the relationship between historical and real time stock market data to see if ML models could predict future prices is important.

Due to the extreme volatility of the equity capital markets, professionals within both the industry and academia have expressed a desire for more research into the topic of stock price prediction. Additionally, as Machine Learning begins to show more and more promise and capabilities, many different experts conduct research that studies the intersection between ML and the financial markets, especially to explore and analyze how ML algorithms can be used to anticipate market trends. For example, research has been done on the variety of different models that can be utilized in order to predict stock prices, as well as the potential obstacles that the industry may face through the use of this technology.

Dataset

This project utilizes a Python module named yfinance. Yfinance allows one to retrieve any company's historical stock prices data from Yahoo. The dataset includes the lowest price, the highest price, the opening and closing prices, dividend date and the stock splits. This model used Apple Inc. (ticker - AAPL) stock price dataset that

contains only the opening, closing, and high price from the last five years. Utilizing 1,250 of the most recent stock prices to predict a future outcome, 80% of these are used to train the algorithm. The remaining 20% are used to test whether the predictions that the algorithm makes are accurate.

Methods/Methodologies

Prior to using any machine learning model, the data had to be preprocessed to extract information such as opening, closing, and high prices. Afterwards, the data was divided into two groups - train and test datasets. Approximately 80% of the data was assigned to the training dataset, while the remaining 20% was assigned to the test dataset. Figure 1 illustrates how the data was partitioned into the two groups. The training dataset was used to train a neural network model, which was then employed to make predictions on the test dataset. The sklearn library was utilized to accomplish this task.
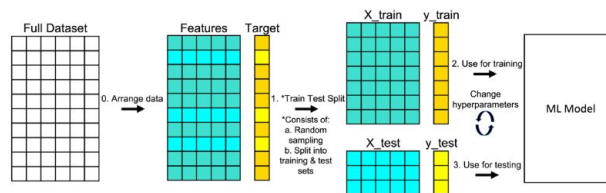
**Figure 1.** Demonstration of Splitting of Data Into Test and Train Groups

The initial model used for regression was a linear model, which is considered to be the simplest form of modeling. It works by creating a line of best fit based on the available data points to predict unknown values. Since the task was to predict a numerical value within a given range, i.e., 1-5, linear regression was considered appropriate. The model utilized the past ten stock prices to predict the opening, closing, and high price for the next day. Although the stock market showed some sharp spikes in both positive and negative directions, the linear model performed well, and achieved a test accuracy of 99.74%. Figure 2 provides an illustration of how linear regression works by creating a line of regression through the entire set of data points.
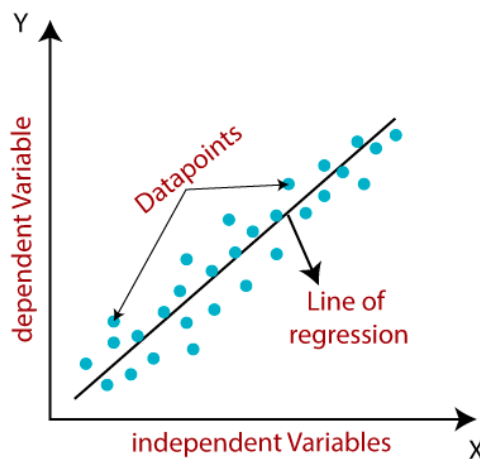
Figure 2. Example of how Linear Regression Functions

Next, the K-nearest neighbors model was applied for classification. This model works on the principle that similar data points are located close to each other, and calculates the distance between them. Figure 3

illustrates how the K-nearest neighbors classification model labels different data by using this distance metric. The model analyzes past data points and identifies patterns that best match current conditions, which enables it to make effective predictions. The K-nearest neighbors model achieved a test accuracy of 99.61%.
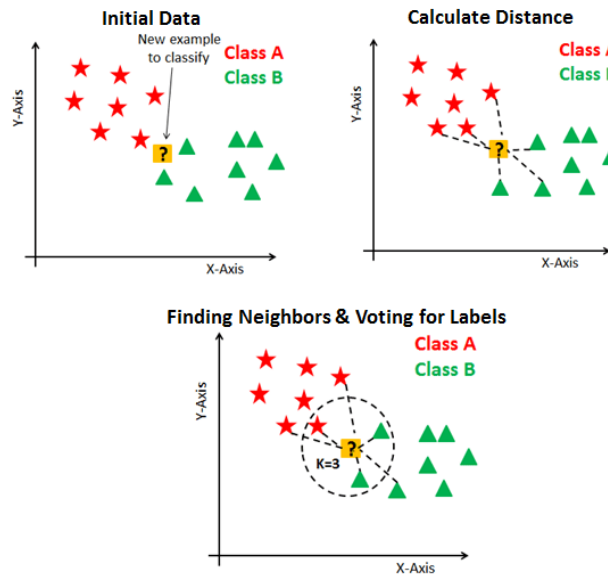
Figure 3. Demonstration of K-Nearest Neighbors Classification Model

The model is trained on available data to build a predictive model, which is fine-tuned through multiple iterations with different inputs to improve its accuracy.

Another option for classification modeling is decision trees, which use tuned parameters to predict results and follow a top-down approach to the dataset during the training phase. Figure 4 illustrates how nodes are used by a decision tree to make predictions. The decision tree model achieved a test accuracy of 99.44%. Decision trees are particularly effective when dealing with datasets that contain fewer classes, such as the case with only two classes in this study (positive correlation and negative correlation).
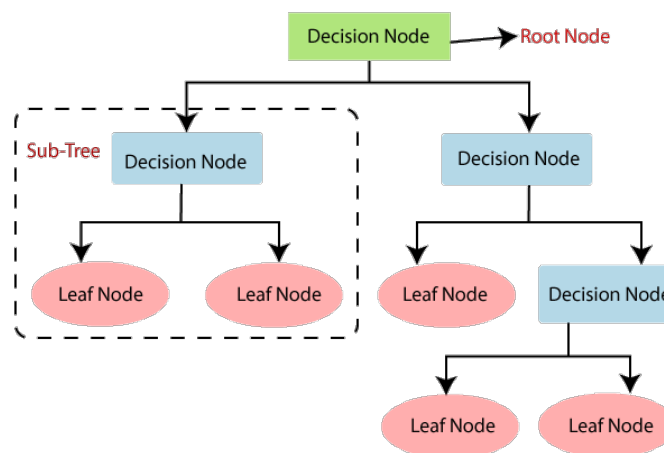
Figure 4. Demonstration of Decision Trees Classification Model

The primary algorithm used in this study was a MLP classifier, which relies on a neural network to

carry out the prediction procedure. A neural network is a computer learning method inspired by the workings of the human brain, utilizing interconnected neurons to recognize complex input questions with minimal human intervention. Figure 5 depicts the multiple input layers and intricacies of the hidden layers that a MLP classifier utilizes to predict the correct response to a stock price prediction question. This model is well-suited to this research problem because it can interpret the relationship between nonlinear and non-associated input and output data. This is particularly important in financial markets, which are highly unpredictable and volatile. As a result, the MLP classifier works optimally to predict responses to stock price prediction.
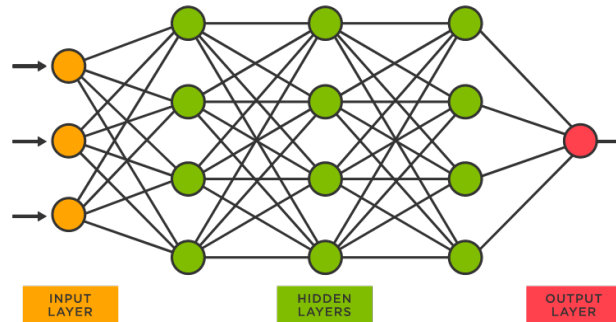

Figure 5. Example of Neural Network with Layers Used

Results

From the research findings, it could be synthesized that using AI models in order to predict responses to stock price prediction is highly successful. This best result was obtained using a MLP classification model (neural networks) which utilized an extremely high number of different inputs in three categories: the high price, opening price, and closing price. In this case, each input was an individual stock price within either a certain range of time - such as the last 5 years - or a set number of previous stock prices - this number generally ranged from 5-10. Using this model, a test accuracy of 99.75% was achieved which means that the model could predict the more accurate price of a stock. However, the x-train accuracy that was received is an error margin of 1.516%.

Table 1. Various Classifier Models and Results

Model	Train Accuracy (Error %)	Test Accuracy (%)
Neural Networks	1.516	99.75
Decision Tree	2.404	99.44
K-nearest neighbors	2.075	99.61
Linear Regression	1.45	99.74

Discussion

It is extremely difficult for both humans and machines to predict stock prices to the exact value, regardless of the amount of data due to the volatile nature of the financial markets, training accuracy was measured in terms of how many predictions were outside a margin of error , which was plus or minus 2 points. When comparing the training and test accuracies, it is to be noted that the close proximity of the percentages indicates that the AI model had no overfitting which is positive. The process of overfitting is when an ML model has been too sensitive to the current dataset it is processing on; therefore, it loses its applicability to other datasets and situations rendering the model unusable.

The accuracy metric was used to compare the results from different models. Table 1. summarizes all of the models used and their respective train and test accuracies. It can be seen that the MLP Regressor (neural network) had the best test accuracy, even though the linear regression had the best train accuracy. In this case, overfitting may have been an issue for the decision tree and k-nearest neighbors classification models.

We used hyperparameters to fine tune our dataset - only within the past few years or a set number of previous data points.. Feeding the model more values would lead to a dataset that is perhaps too broad and wouldn't accurately reflect the current market trends and conditions.

Conclusion

This study explored the efficiency of machine learning algorithms in predicting stock prices. The results indicate that the MLP classifier neural network and linear regression model, in particular, are capable of producing accurate predictions when trained with relevant financial data. The findings of this study suggest that incorporating supervised machine learning techniques in the stock market can enhance the decision-making process for investors. However, it is important to note that the stock market is inherently volatile and unpredictable, and no algorithm can provide a foolproof prediction of future prices. Thus, investors should use these predictions as a tool for informed decision-making rather than relying on them entirely.

Future research in the field of stock price prediction can be extended to other neural network models, such as CNN, RNN, and LSTM models to determine more accurate predictions. One potential direction is the exploration of additional data sources that may have an impact on stock prices. For example, macroeconomic indicators such as inflation rates, GDP growth, and interest rates could potentially be incorporated into models to provide a more comprehensive understanding of the market. Overall, there are many exciting opportunities for future research in the field of stock price prediction, and continued exploration of these areas can lead to significant advancements in the field.

Acknowledgments

References

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS One*, *12*(7).

Daori, H., Alharthi, M., Alanazi, A., & Alzahrani, G. (n.d.). Predicting Stock Prices Using the Random Forest Classifier. *Research Square*.

International Systems in Accounting, Finance and Management, *26*(4). (2019).

K. C, S. A. N. A, A. P, P. R, M. D and V. B. D, "Predicting Stock Prices Using Machine Learning Techniques," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 1-5, doi: 10.1109/ICICT50816.2021.9358537.

Mehtab, S., & Sen, J. (n.d.). *A Time Series Analysis-Based Stock Price Prediction Using Machine Learning and Deep Learning Models*. https://arxiv.org/pdf/2004.11697.pdf

Sahoo, P., Dr., & Charlapally, K. (2015). Stock Price Prediction Using Regression Analysis. *International Journal of Scientific & Engineering Research*, *6*(3).

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock Closing Price Prediction using Machine Learning Techniques. *Science Direct*, *167*.