

Applying AI Models for Stock Investment Decisions

Diya Shiburam

Department of Computer Science, Cornell University, New York, USA

Email: od84@cornell.edu

How to cite this paper: Shiburam, D. (2024) Applying AI Models for Stock Investment Decisions. *Open Journal of Applied Sciences*, 14, 3061-3068.

<https://doi.org/10.4236/ojapps.2024.1411201>

Received: May 20, 2024

Accepted: November 10, 2024

Published: November 13, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In the realm of finance and economics, the search for reliable stock predictions remains a focal point for researchers [1]. Contrary to common belief, stock market prices are not merely random guesses, but rather powerful tools that allow people to decide where to put their money, proving to be a significant aspect of finance. In this paper, by applying machine learning techniques, we propose to predict stock prices based on trends from previous years' stock data using learning models, such as Linear Regression, MLP Regressor, Decision Tree Regressor and Random Forest Regressor. To enhance the model's decision-making capabilities, the model was programmed to decide whether to sell or buy stocks using the predictions from the linear model. If the model anticipates an increase in stock prices, it suggests buying more stocks. On the contrary, if the model predicts a downturn, it suggests selling stocks in order to benefit the investor and enhance profitability. If the investor began with no stocks and \$20,000, through the use of our model, the investor was able to make 161.3% profit. In another scenario where the investor holds 200 stocks and \$10,000, the investor was able to make a 546.3% profit. Ultimately, the model results in profitable outcomes.

Keywords

Stock Market, Trading, Artificial Intelligence, Finance

1. Introduction

The stock market aids with the buying and selling of shares from publicly traded companies. The shares (or stocks) represent ownership within a company. By purchasing and selling these shares, investors engage in the stock market, giving companies the chance to grow and giving investors the chance to profit from the success of companies in a variety of industries. This process is shaped by not only

economic factors, but also social and political influences [2].

Dating back to 13th Century Europe, where merchants and traders gathered in marketplaces to buy/sell commodities, debts or shares, these early markets laid the groundwork for establishing more formalized stock exchanges. The concept of companies issuing shares to raise capital gained popularity during the 18th Century, particularly in the United States and continued to develop throughout the 19th and 20th Centuries, alongside new electronic trading systems that increased market efficiency and accessibility. A majority of modern-day trading is driven by algorithms that use complex mathematical models to execute trades at more efficient rates. Overall, the stock market holds immense importance and significance due to its profound impact on the global economy and financial landscape.

Predicting stock market prices has been an objective for investors for decades. The appeal of being able to foresee the movements of financial markets comes from the potential for profits and wanting to mitigate economic risks [3]. As new and innovative technology emerges, the search for more accurate predictions continues, highlighting the impact of technology within finance and market analysis.

Through this paper, we aim to forecast stock prices, attempting to make accurate predictions about future prices based on historical data. This provides investors with valuable information about potential market trends, improving their ability to optimize variables and forecast future prices for ideal outcomes. With accurate predictions, investors can make well-informed decisions about when to buy or sell stocks, leading to better profits and portfolio management.

2. Dataset

The first step of building an AI model is to have access to an accurate and reliable dataset. Utilizing the “yfinance” library, the historical data is organized into a pandas Data frame, focusing solely on the opening prices from the last 5 years to focus our predictions. The library allows us to specify the time frame, enabling us to adjust the number of years of data we want to employ for our analysis. Datasets serve as a training ground for predictive modelling, with recent studies highlighting the role of generative AI models in improving data analysis [4]. By feeding historical data into machine learning algorithms, it allows the models to form predictions on new or unseen data, that can be used to develop predictions for future stock prices. After gathering the data, the dataset is typically split into training and testing sets.

The training set is used to train the model on historical data, allowing the model to interpret patterns and relationships between the input features and target variables, while the testing set is used to evaluate the model. Within the model, 1/3 of the data is saved as the testing set, while the remaining is utilized to train the model. The dataset isn't limited to only open prices, it can be applied to a wider range of stock markets, including dividends and stock splits. Essentially, the dataset acts as the foundation for the predictive modelling system, providing the model with precise data that allows it to analyze trends to predict future prices.

3. Explaining the Models

In our research, the first model we employed as our initial approach for predicting stock prices is a linear regression model. A linear model is a learning algorithm that works by predicting the relationship between a dependent variable and independent variables by fitting a linear equation. Using the formula “ $P = w_1p_1 + w_2p_2 + w_3p_3$ ”, where each parameter is multiplied with the corresponding weight, a linear model predicts by adjusting these weights to produce the most accurate predictions. In this case, p_1 , p_2 , and p_3 represent the last three open prices. This number can be adjusted to include more open prices if needed. The model does not require large amounts of data as it has a relatively small number of parameters.

The second model we employed is a Decision Tree Regressor. Unlike linear models, which approximate weights through linear combinations, decision tree models work by applying a tree-like model of decisions and their subsequent consequences to make predictions. Inspired by the human decision-making process, the model evaluates options based on a set of criteria and chooses the option that best fits the objective, as shown in **Figure 1**. The decision tree typically starts with one node that branches off into various possible outcomes.

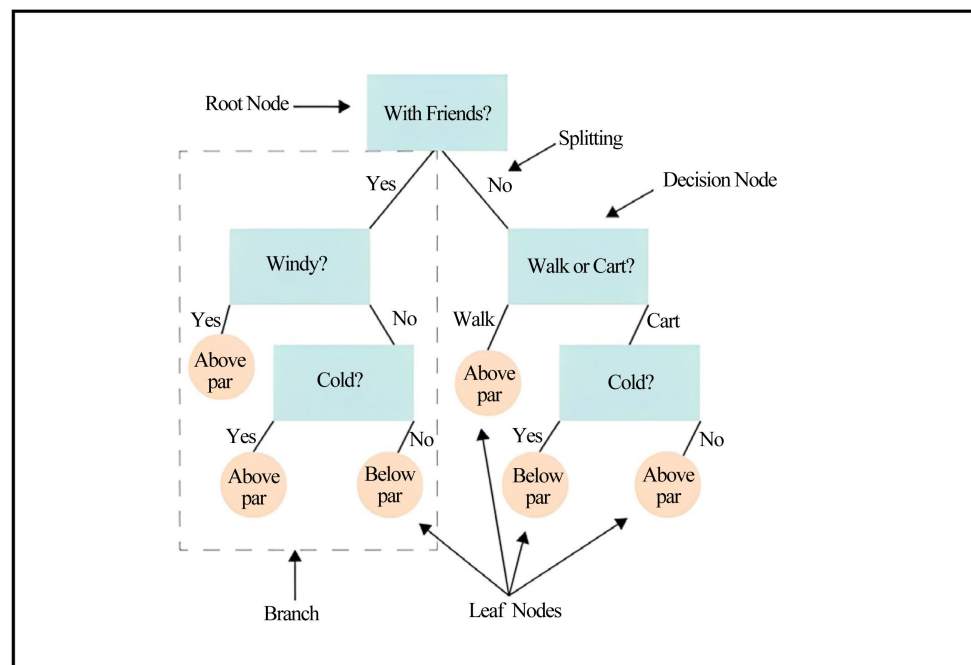


Figure 1. Example of a decision tree regressor process.

The third model we utilized was an MLP Regressor. “MLP” stands for Multi-Layer Perceptron, a type of artificial neural network made up of multiple layers of nodes, or neurons. Unlike simpler models like linear regression, the MLP can handle more complex relationships between inputs and outputs. It does this by using layers of neurons to understand and capture non-linear relationships between the

input and output variables, as shown in **Figure 2**. While both models adjust weights to make better predictions, the MLP can handle more complicated patterns in the data.

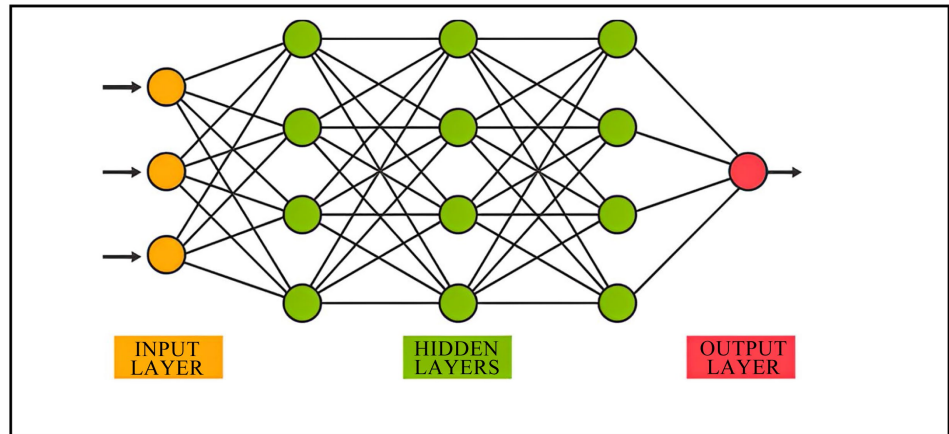


Figure 2. Example of a MLP Regressor process.

The final model we tested was a Random Forest model. Unlike the Decision Tree Regressor, the Random Forest model employs multiple decision trees. Instead of relying on the prediction of a single tree, it combines the predictions from numerous trees by averaging their outcomes, allowing it to handle more complex datasets, as shown in **Figure 3**.

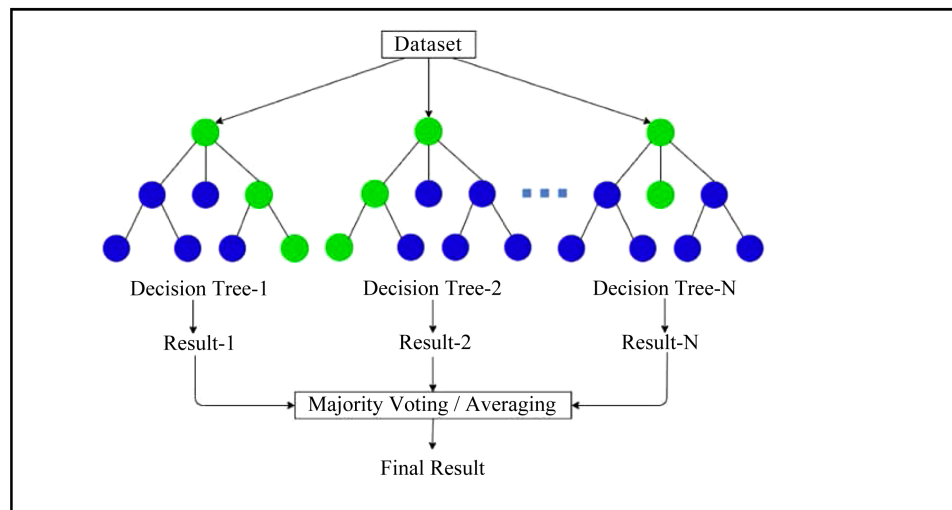


Figure 3. Example of a Random Forest model process.

Finally, in order to evaluate the accuracy and efficiency of the models, we use MSE (mean squared error). The MSE measures the number of errors in statistical models by subtracting the predictions from the outcomes, squaring the results and then calculating the average, as shown in **Figure 4**. A larger MSE indicates there is more error, while a smaller MSE suggests the opposite.

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

Figure 4. Mean squared error equation.

Table 1. MSE of each model on different companies (on the test set).

| | Linear Regression | Decision Tree | Random Forest | MLP Regressor |
|-----------|-------------------|---------------|---------------|---------------|
| Apple | 6.64 | 13.41 | 8.57 | 7.51 |
| McDonalds | 8.88 | 16.55 | 10.86 | 10.82 |
| Uber | 1.60 | 3.32 | 2.05 | 1.64 |
| Nike | 5.91 | 11.48 | 6.86 | 6.53 |
| Amazon | 8.89 | 17.73 | 10.92 | 9.74 |

To ensure that the model doesn't overfit, we can compare the testing error (**Table 1**) to the training error (**Table 2**). A significant difference between the training error and MSE (testing error) suggests overfitting, as it indicates that the model performs well on the training data but poorly on new, unseen data. On the other hand, if the MSE on the test data remains similar to the training error, it suggests that the model has learned the general patterns in the data rather than memorizing the training set. This indicates that the model has not overfit the dataset.

Table 2. Training error of each model on different companies.

| | Linear Regression | Decision Tree | Random Forest | MLP Regressor |
|-----------|-------------------|---------------|---------------|---------------|
| Apple | 7.72 | 0.0 | 1.45 | 8.39 |
| McDonalds | 8.45 | 0.0 | 1.57 | 10.85 |
| Uber | 1.79 | 0.0 | 0.29 | 1.88 |
| Nike | 5.48 | 0.0 | 0.91 | 6.19 |
| Amazon | 9.81 | 0.0 | 1.57 | 12.14 |

When comparing the MSE values to the training error values, displayed in **Table 2**, across different models, it is observed that the linear regression and MLP Regressor models exhibit training error values that are very similar to their respective MSE values. In contrast, the Decision Tree Regressor and Random Forest Regressor show a noticeable variation between their training error and MSE values. This suggests that while the linear models demonstrate more consistency between their training performance and how well they adapt to new data, the tree-based models may exhibit overfitting, leading to larger inconsistencies between their training and testing errors. To ensure that the model can effectively generalize to different companies despite market volatility, we included examples from a variety of companies across distinct industries and sectors, highlighting the model's ability to adapt to varying market conditions.

4. Explaining the Simulation

To leverage the models, we created a simulation that predicts the potential profit from buying or selling stocks. We chose to employ the linear regression model as it consistently achieved a lower mean squared error (MSE) compared to the other models. This is because, with a limited amount of data, simpler models tend to perform better in comparison to complex models. However, if we manage to obtain more data, more complex models may perform better than the linear regression model.

To achieve this, we implemented a straightforward algorithm. If the predicted price is higher than 2 points compared to the current value and there is sufficient budget, the algorithm buys 2 stocks and deducts the cost from the budget. Additionally, if the predicted price is lower than 2 points compared to the current price and there are enough stocks, it sells 2 stocks and adds the earnings to the budget. For smaller predicted price changes, the algorithm buys 1 stock if the predicted price is higher than 1 point and there is enough budget or sells 1 stock if the predicted price is lower than 1 point and at least 1 stock is owned. This is a simple strategy that can be adjusted to fit more aggressive stock purchasing during predicted increases and more aggressive selling during predicted decreases. This approach ensures that the simulation is adapted to both significant and minor market fluctuations.

While the model does not incorporate a formal risk management strategy, it is designed to execute buy or sell decisions only when the predicted stock price deviates significantly from the current market price. This ensures that actions are only taken when the model has strong confidence in a large price change. Through the simulation, we also account for market liquidity by ensuring that there is sufficient budget available before executing any stock transactions. Additionally, we incorporate a transaction fee of \$0.10 for each buy or sell order to reflect real-world trading costs [5].

5. Results

The results of the simulation showed significant variations in profitability among different companies, highlighting how the method performs differently depending on the stock. As shown in **Table 3** and **Table 4**, Apple showed the biggest percentage profit, demonstrating a good match between the model's forecasts and the actual performance of the stock on the market. In contrast, Nike and Uber displayed relatively small increases in profit (specifically in **Table 3**), indicating that the model was less effective in forecasting stock price movements. Amazon and McDonald's profits were within the average range, with Amazon making far more profit than McDonald's, likely due to the model having a better match with Amazon's market trends. It is important to recognize that the profit is not only based on the simulation, but also on the general progress or volatility of the stock market.

Table 3. Results of simulation (starting budget = \$10,000).

| Company | Beginning Budget | Final Sum | Percentage Profit |
|-----------|------------------|------------|-------------------|
| Apple | \$10,000 | \$29798.93 | 197.99% |
| McDonalds | \$10,000 | \$14698.53 | 46.99% |
| Uber | \$10,000 | \$14194.38 | 41.94% |
| Nike | \$10,000 | \$13138.40 | 31.38% |
| Amazon | \$10,000 | \$18651.70 | 86.52% |

Table 4. Results of simulation (starting budget = \$2000).

| Company | Beginning Budget | Final Sum | Percentage Profit |
|-----------|------------------|-----------|-------------------|
| Apple | \$2000 | \$7104.59 | 255.23% |
| McDonalds | \$2000 | \$3036.69 | 51.83% |
| Uber | \$2000 | \$4400.49 | 120.02% |
| Nike | \$2000 | \$3357.80 | 67.89% |
| Amazon | \$2000 | \$3908.18 | 95.41% |

6. Open Questions

Analyzing the outcomes of our simulation and the varied profitability shown by the different stocks prompts us to think about several insightful questions. How can we use financial reports in our data to refine the predictions? How do changes in the news, social media, political changes and major conflicts affect stock prices and how can we incorporate this into our predictions?

7. Conclusion

In conclusion, using AI algorithms for stock price prediction and investment decision-making can transform the way individuals view stock market investing. Through the usage of machine learning techniques like Random Forest Regressor, MLP Regressor, Decision Tree Regressor, and Linear Regression, investors can make well-informed decisions based on past data trends. For instance, employing the Linear Regression model within the simulation resulted in an average percentage increase in profit of 99.52%. Furthermore, combining real-time data from news, social media, and financial reports can improve predictions, giving investors the ability to anticipate market trends and make more informed investment decisions. Ultimately, using AI in stock market investment can shape the future of the stock industry.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Li, T. (2024) Stock Prediction and Analysis Based on Machine Learning Algorithms.

Applied and Computational Engineering, **50**, 15-22.

<https://doi.org/10.54254/2755-2721/50/20241142>

- [2] Mollema, W.J.T. (2024) Decolonial AI as Disclosure. *Open Journal of Social Sciences*, **12**, 574-603. <https://doi.org/10.4236/jss.2024.122032>
- [3] Mokhtari, S., Yen, K.K. and Liu, J. (2021) Effectiveness of Artificial Intelligence in Stock Market Prediction Based on Machine Learning. *International Journal of Computer Applications*, **183**, 1-8. <https://doi.org/10.5120/ijca2021921347>
- [4] Ratican, J., Hutson, J. and Wright, A. (2023) A Proposed Meta-Reality Immersive Development Pipeline: Generative AI Models and Extended Reality (XR) Content for the Metaverse. *Journal of Intelligent Learning Systems and Applications*, **15**, 24-35. <https://doi.org/10.4236/jilsa.2023.151002>
- [5] Liu, Z., Mu, C. and Wen, C. (2016) Impacts of Internal Financing on Investment Decisions by Managers with Cognition Biases. *Journal of Mathematical Finance*, **6**, 431-442. <https://doi.org/10.4236/jmf.2016.63034>