Research Article

Leveraging AI Machine Learning Models to Predict and Enhance Long-Term Investment Efficiency

Pranav Prasath

Lambert High School, 85 Highgrove Drive, Suwanee, GA, USA

ABSTRACT

The stock market is a very tedious, difficult, and rewarding area that many people all across the world use to invest. However, there are some issues. As time goes on, the stock market becomes forever increasing and, in return, more intricate and complex to figure out. Getting to know the markets takes long hours of researching and analyzing data from the past and applying it to modern-day norms. This is why only certain people are suited for this type of work. Accurate stock price predictions benefit investors and traders in many ways. It helps with increased efficiency in the trading hours and creates higher margins and profits for these people. By using accurate predictions, they can further be more confident in purchasing and selling, which overall makes their job easier and more positive than what is usually negative. The purpose of this study is to showcase AI being applied for economic and financial enhancements such as investing. As a team, we have found that using a bot that predicts, buys, and sells stocks greatly increases investments. We used a select amount of machine learning models that were trained and tested to get the most accurate predictions possible to use in our Bot. We found that using an accurate model increases profits for investments in different brands many times over. This implies that our research and methods work well and are not faulty, as they were tested across many different companies at unique amounts.

Keywords: Stock Market, Prediction model, AI Model, Investing, Machine Learning

INTRODUCTION

Stock price prediction has long been a subject of interest for researchers, traders, and financial analysts due to its potential to yield significant financial returns. The dynamic nature of stock markets, influenced by a variety of factors including economic indicators, market appeal, and geopolitical events, makes accurate prediction a challenging task. With the rise of machine learning and artificial intelligence, new methods have emerged that offer more sophisticated approaches to predicting stock prices, leveraging patterns and relationships within the data that traditional models may overlook.

In this project, we focused on developing machine learning models to predict stock prices using historical price data and market indicators. The goal is to enhance the accuracy of predictions by employing a diverse set of machine-learning techniques, each with its own strengths. Specifically, we implement four models: Linear Regression, Multi-layer Perceptron (MLP) Regressor,

Corresponding author: Pranav Prasath, E-mail: pranavprasath09@gmail.com.

Received February 17, 2025; **Accepted** March 17, 2025
https://doi.org/10.70251/HYJR2348.32912

Decision Tree Regressor, and Random Forest Regressor. These models were chosen for their ability to capture both linear and non-linear relationships, allowing for a more comprehensive analysis of stock price movements.

In addition to model performance, the success of a stock price prediction system is often measured by its ability to inform profitable trading strategies. To this end, two trading bots were developed that simulate buying and selling stocks based on the predictions generated by the machine learning models. The first bot operates on an explicit strategy, buying or selling one stock per trade based on the predicted price movement for the next day. The second bot adopts a more aggressive strategy, trading two stocks per transaction, which increases the potential for both gains and losses.

The performance of the models is evaluated using the Mean Squared Error (MSE), which assesses the average deviation between the predicted and actual stock prices. This serves as a foundation for comparing the predictive power of the different models, while the two trading bots provide insight into how well these predictions can be translated into profitable trading decisions. By combining machine learning models with automated trading strategies, this project aims to contribute to the ongoing exploration of how AI can be used to enhance financial decision-making (1-4).

MATERIALS AND METHODS

In this section, we describe the models and methods used for stock price prediction and the trading strategies that were developed based on these predictions. Four machine learning models were employed for predicting stock prices: Linear Regression, MLP regression, Decision Tree Regressor, and Random Forest Regressor. We also evaluated the models based on their performance using the Mean Squared Error measurement. Additionally, two trading bots were created to simulate trading strategies based on the predictions.

Machine Learning Models

1. **Linear Regression**:
   Linear regression is a very simple yet effective model used to establish a relationship between the independent and the dependent variable. In our case, the target is the stock price, and the features include past prices and other relevant market indicators. Linear regression attempts to fit a straight line that minimizes the error between the predicted values and the actual stock prices, which is called a line of best fit.
2. **MLP Regression**:
   The MLP regression is a type of artificial neural network that consists of multiple layers of neurons, including input and output layers. This model is more complex than linear regression and has the ability to find nonlinear relationships in the data. In return, it may give a better fit depending on the data given. The MLP regressor is particularly useful when dealing with time series data like stock prices, where complex patterns might exist.
3. **Decision Tree Regressor**:
   Decision trees use a tree-like structure to split the data into smaller segments based on specific values. At each node in the tree, the model chooses the feature and value that results in the best split, minimizing the prediction error for each segment. The decision tree regressor can find non-linear patterns in the data and provides a simple, interpretable structure for making predictions.
4. **Random Forest Regressor**:
   The random forest regressor is an ensemble learning method that builds multiple decision trees and combines their predictions. Each tree is trained on a different subset of the data, and the final prediction is the average of the predictions made by all the trees. Random forests are known for their robustness and ability to generalize well on missing data, which diminishes the risks of overfitting.

Mean Squared Error (MSE)

The performance of the models was evaluated using the ***Mean Squared Error (MSE)*** metric. MSE measures the average squared difference between the actual and predicted stock prices. This metric provides a sense of how close the model's predictions are to the true values. It is calculated as:

MSE=n1i=1∑n(yi−yi^)^2

Where:

- n is the number of data points
- yi is the actual stock price
- yi^ is the predicted stock price

A lower MSE indicates better model performance, as it means the predictions are closer to the actual values.

Trading Bots

To simulate trading strategies based on the stock price predictions, two trading bots were developed:

1. **Trading Bot 1**:
   This bot buys or sells ***one stock*** based on the predicted price movement for the next trading day. If the model predicts that the stock price will rise, the bot executes a buy order. Alternatively, if the model predicts a price drop, the bot executes a sell order. This approach is a simple one-stock-per-transaction strategy aimed at leveraging short-term price movements.
2. **Trading Bot 2**:
   In contrast to the first bot, ***Trading Bot 2*** buys or sells ***two stocks*** based on the predicted price movement. This increases the exposure to potential gains or losses. The bot follows the same logic as Trading Bot 1: a predicted increase in stock price triggers a buy order, while a predicted decrease triggers a sell order. By trading two stocks instead of one, the bot takes on a slightly higher level of risk, but with the potential for larger profits if the model's predictions are accurate.

Both bots are designed to make trades based on daily predictions, simulating a short-term trading strategy where decisions are made for the next day based on the stock price forecast. The performance of each bot was evaluated to compare the profitability and risk associated with different trading volumes.

RESULTS

Tables and statistics of models

The Trading Bot One shows a simulation of how well the models would work if we gave it a certain amount to invest and what the final amount is (Table 1). This Trading Bot can only sell and buy one stock at a time.

The Trading Bot Two shows a simulation of how well the models would work if we gave it a certain amount to invest and what the final amount is (Table 2). This Trading Bot can buy and sell two stocks at a time instead of just one.

The Training Mean Squared data table shows the mean squared error for each model on the training sets data (Table 3).

The Testing Mean Squared data table shows the mean squared error for each model on the testing sets data (Table 4).

DISCUSSION

The findings of this study show that through proper training on past stock data, machine learning models can offer useful information for predicting stock prices. The Random Forest model consistently performed better in predicting accuracy compared to the Linear Regression, MLP Regressor, and Decision Tree models, as shown by its lower Mean Squared Error (MSE). This implies that using powerful machine learning models such as Random Forest, which merges various decision trees, is especially effective for detecting intricate patterns in financial information.

In the world of errors, people deserve to understand what they mean and what the cause and effects of them

Table 1. Trading Bot 1

Trading Bot	Google	Apple	Microsoft	Tesla	Nike
$4000	$11037	$14349	$11938	$50825	$6222
$3000	$8864	$11046	$9020	$42959	$4865
$2000	$5754.0	$7630	$6088	$36988	$3583
$1000	$2715	$3932	$2848	$24755	$1718

Table 2. Trading Bot 2

Trading Bot	Google	Apple	Microsoft	Tesla	Nike
$4000	$10115	$15338	$11818	$76163	$7166
$3000	$7630	$12132	$8836	$64971	$5196
$2000	$5093.0	$8088	$5476	$51431	$3437
$1000	$2553	$3970	$2529	$35171	$1626

Table 3. Training Mean Squared Table

Training The Data	Google	Apple	Microsoft	Tesla	Nike
Linear Regression	6.49	8.75	22.09	73.96	5.75
MLP	6.67	10.47	26.06	71.81	6.78
Decision Tree	11.7	15.5	46.79	142.93	0
Random Forest	7.61	10.66	28.69	92.91	1

Table 4. Testing Mean Squared Table

Test The Data	Google	Apple	Microsoft	Tesla	Nike
Linear Regression	5.14	7.83	20.66	72.81	5.12
MLP	5.51	9.92	25.8	75.88	6.28
Decision Tree	13.15	19.78	48.07	193.83	10.08
Random Forest	0.93	1.35	3.76	12.7	6.19

encompass. The Mean Squared Error is a highly effective method that aims to determine how accurate a Machine Learning model is at predicting data after being trained with previous or similar data. Different models, however, are best suited for different purposes depending on what they do with training data. For example, in this case the Decision Tree model showed a substantially higher measure (MSE) due to its much more abstract and simple approach comparatively to its complex relative, the Random Forest Model. A higher measure of error indicates a less accurate prediction, which is not optimal for this study. Similarly, the Linear and MLP Regressions exhausted a higher error somewhere in between the Random Forest and Decision, concluding that the Random Forest model is the best in this specific situational study.

Understanding the power of these AI models and bots is crucial, as there are limitations as to what is possible in this era. Limitations include dictating when significant events occur, such as geopolitical conflicts and economic sanctions throughout the world. AI models that predict on new data while using older data to train may not be able to respond and react to these events that may drastically upshift or downshift the overall structure of the stock market. This is a disadvantage for this level of stock prediction, which creates a need for future expansion ideas in the near future as AI and technology are ever-expanding facets of life. Introducing late-stage reinforcement learning to train the models on the modern consensus would successfully incorporate this idea.

Additionally, the effective use of these predictions with our two trading bots demonstrates the opportunity for making money in an actual trading setting. Both bots, which determined their trading decisions by predicting stock price movements, were successful in generating profits during the simulated trading period. Trading Bot 1, which used a conservative approach of purchasing or selling one stock at a time, showed consistent and gradual progress. This method was successful in reducing risk while still utilizing the model's predictive capability.

Trading Bot 2, which has a more dominant approach of purchasing or selling two stocks per trade, resulted in increased profits overall but also brought about higher fluctuations in the overall performance. This indicates that although aggressive strategies can boost profits when predictions are correct, they run the risk of higher losses in times of market uncertainty. However, the continual profitability in both bots shows that incorporating machine learning forecasts into trading approaches can result in positive results.

The implementation of these bots would require a significant amount of capital investment and marketing techniques to propel itself into the modern market. Due to the fact that these bots are almost fully autonomous, people would most likely have difficulty understanding it and building up trust towards these newer technological methods. Additionally, a simulation would be beneficial for customers to see if the bots would work for them. This process could take a month or more and could provide more information to consumers. This would call for vigorous amounts of testing, proving, and advertising to markets as a sincere and successful way to start a particular journey in the vast world of stocks. To commit to all of these factors, it would take an intelligent team of people who are dedicated to the goal and a certain sum of money to get started researching for bigger, long-term plans.

To sum up, this project shows that machine learning models, specifically Random Forest, are successful in forecasting short-term fluctuations in stock prices. Through the use of these predictions with automated trading techniques, we have successfully demonstrated the ability to produce returns on a computer-simulated trading platform. Potential future research could be further investigated for additional improvements to the models.

ACKNOWLEDGEMENTS

I would like to thank Odysseas Drosis from Cornell University for their guidance on this project.

DECLARATION OF CONFLICT OF INTERESTS

The author(s) declare that there are no conflicts of interest regarding the publication of this article.

REFERENCES

1. Qin Y, Huang Y, Lu L, Wang Z. The Application of Machine Learning to Business Process Management: A Literature Review. *Procedia Comput Sci*. 2020; 176: 687-694.
2. Wang A, Unger J, Ma C, Rong H, Duan X, Jin Z, Liu Y. A Survey on Open-Domain Question Answering. arXiv, 2111.01137, 2021.
3. Hu YC, Patel M, Sabella D, Sprecher N, Young V. Mobile Edge Computing-A Key Technology Towards 5G. *IEEE Trans Commun Mag*. 2015; 52 (5): 126-131.
4. Koivisto J, Oksanen A. Sociability and Spectatorship in Hybrid Virtual Spaces. *Humanit Soc Sci Commun*. 2024; 11: 408.