

Predicting Google Stock Prices Using Machine Learning Models

Zixiong Zhao*

Department of Economics, Syracuse University, New York, United States

*Corresponding author: zzhao101@syr.edu

Abstract. Stock price prediction could help investors decide to sell or buy the stock, which returns and minimizes risk. Precise forecasts can bring substantial financial gains, while unprecise forecasts can bring significant financial losses. During the volatile financial markets, a reliable stock price prediction is a valuable source for traders, investors, and financial analysts. The research uses different models to compare and analyze Google's stock price prediction. The research uses machine learning models to predict Google's stock price. The researcher aims to find the best model to predict the stock price using the Mean Square Error (MSE) result. The researcher chose Google as the topic because of Google's importance in technology and the stock market. The research uses Google's stock data from 2016-2021. The data includes opening, closing, and high, and low prices. The study used three methods, linear regression, decision tree, and random forest, to predict Google's stock market price.

Keywords: Google stock prediction, machine learning models, linear regression.

1. Introduction

Accurately predicting the price of Google stock is critical for investors and market participants and can result in significant financial gains. Stock price forecasting has always been of great interest in the financial markets, as the ability to forecast accurately is critical to risk management. Stock price volatility is influenced by many factors, including global economic conditions and significant events, making it a complex challenge in finance. However, despite the growing importance of stock price forecasting, there has been a lack of comprehensive comparative studies of machine learning approaches to Google stock price forecasting. In addition, the performance differences of different machine learning models have not been studied in depth for historical data from 2016 to 2021. Research on forecasting stock prices has been ongoing. It is yet unknown, though, what kind of stock data is best suited for deep learning, and there is no proof that the generated stock prediction model can turn a profit, as Yoojeong Song wrote in "Importance of Event Binary Features in Stock Price Prediction [1]."

To close this research gap, this study makes use of random forests, decision trees, and linear regression. This study compares and analyses the forecast performance of Google stock using multi-factor data, including starting price, closing price, high price, and low price, utilizing machine learning models including linear regression, decision trees, and random forests. Focusing on predicting the closing price of Google stock, the effectiveness of different machine learning models in capturing the trend and volatility of Google stock price is explored through in-depth analysis of historical data. The ultimate goal is to determine which machine learning model is most effective in capturing the changing trends of Google stock, thus providing investors with a more reliable basis for decision-making. The paper will infer the optimal model for predicting Google's stock price by evaluating the mean square error of model performance, thus filling the knowledge gap in related research areas.

2. Literature Review

Recently, the attempt to forecast stock prices has shifted toward refined machine learning techniques from traditional statistical methods. This shift is predominantly motivated by the

recognition that the intricacies of the stock market's dynamics exceed the capacity of conventional models. Numerous machine learning techniques have been employed in the literature to predict the trajectory of stock prices write in Stock Market Prediction with Gaussian Naïve Bayes Machine Learning Algorithm, by Ampomah, Ernest Kwame. In stock market forecasting, machine learning (ML) has proven to be a particularly effective tool, outperforming conventional fundamental and technical research methods [2].

The stock market can be volatile due to various factors, including global events and the actions of individual investors. Such variables often make predicting future outcomes challenging. Machine learning models, however, can outperform other methods in processing and learning from extensive data sets. Analyzing historical data allows these models to identify underlying patterns and relationships more efficiently than conventional analyses.

Various techniques, such as Artificial Neural Networks (ANN), Random Forests, and Long Short-Term Memory (LSTM) networks, have been utilized to address different aspects of forecasting in the stock market [3, 4]. Financial modeling has diverse applications, from predicting stock prices to identifying prospective opportunities in specific markets. The study presents a comprehensive survey of articles from 2011 to 2020 that applied fusion techniques in various stock market applications. (Ankit Thakkar, Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions) [5]. Furthermore, blending different types of data, such as financial news and conventional stock indicators, can enhance the accuracy of these models.

Therefore, it has been demonstrated that machine learning models can effectively adjust to market fluctuations, and training on diverse data sets can cover the multifaceted nature of financial markets. Their adaptability yields dependable and strong projections of stock market trends, furnishing investors and traders with valuable insights to formulate informed decisions.

3. Methodology and Result

As part of the study, the research analyzed Google's historical stock price data by performing initial data analysis, visualization, and preprocessing. To accomplish this, this paper collected the dataset (GOOG.csv) containing the historical stock prices and conducted an initial analysis of its shape, information, and null values. The research also plotted the opening, closing, and high and low prices to visualize the stock trend.

The research performed features engineering based on closing prices to incorporate the temporal aspect of stock prices. This involved creating a lagged dataset wherein shifted the 'close' prices to create features representing the previous day's stock price.

The researcher trained and tested three models to select the best model in the research: Linear Regression, Decision Tree, and Random Forest. The paper evaluated the performance of the models using Mean Squared Error (MSE).

Linear Regression: A baseline model for regression tasks (Fig. 1) [6]. This model predicts the stock price as linear. It's typically used for its simplicity and efficiency in predicting linear relationships. The independent variable X could be various factors like past prices, volume, or other financial indicators. The model aims to find the best-fitting linear relationship between these variables and the stock price

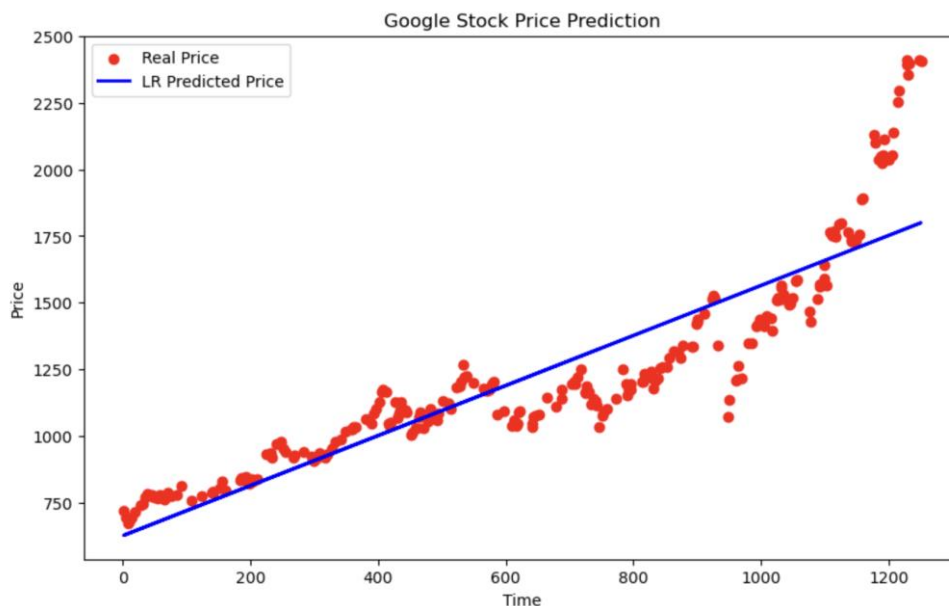


Fig. 1 Linear Regression Prediction (Photo/Picture credit: Original).

Decision Trees: A non-linear model to capture complex patterns (Fig. 2) [7]. This model uses a tree-like graph of decisions and their possible outcomes. It helps capture non-linear relationships that linear models may miss. A decision tree splits the dataset into branches based on input variable values to isolate regions where stock prices behave similarly. The final prediction is determined by calculating the average of the target variable, stock price, in the leaf nodes where the input data falls.

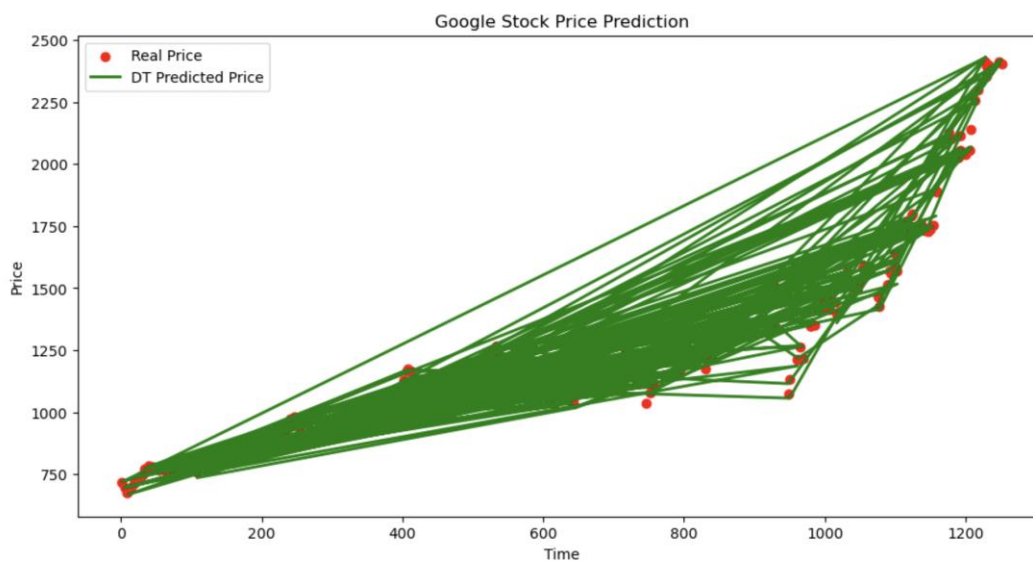


Fig. 2 Decision Trees Prediction (Photo/Picture credit: Original).

Random Forest Regressor: An ensemble model to improve prediction robustness (Fig. 3) [8]. In an ensemble learning method, random forests create a "forest" of decision trees and merge their outputs. The approach is known for its higher accuracy and ability to reduce overfitting. It generates numerous Decision Trees on diverse dataset subsets and merges their predictions. The mean of all tree predictions usually forms the final prediction, which minimizes the common variance and overfitting problems found in single Decision Trees.

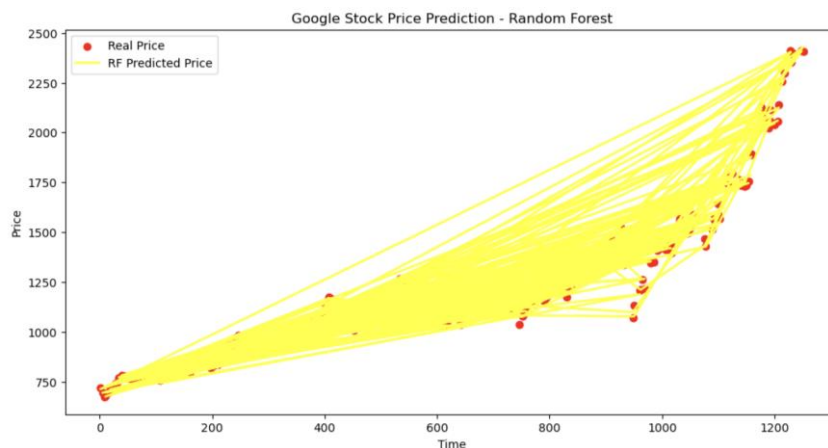


Fig. 3 Random Forest Prediction (Photo/Picture credit: Original).

Learning Curves: The researcher can plot learning curves for each model to understand how performance improves with more training data (Fig. 4, 5, 6) [9]. It will help the researcher visualize how well each model learns from the training data and how they perform in cross-validation. This can provide insight into whether the model is underfitting or overfitting or if more information can be gained from additional data.

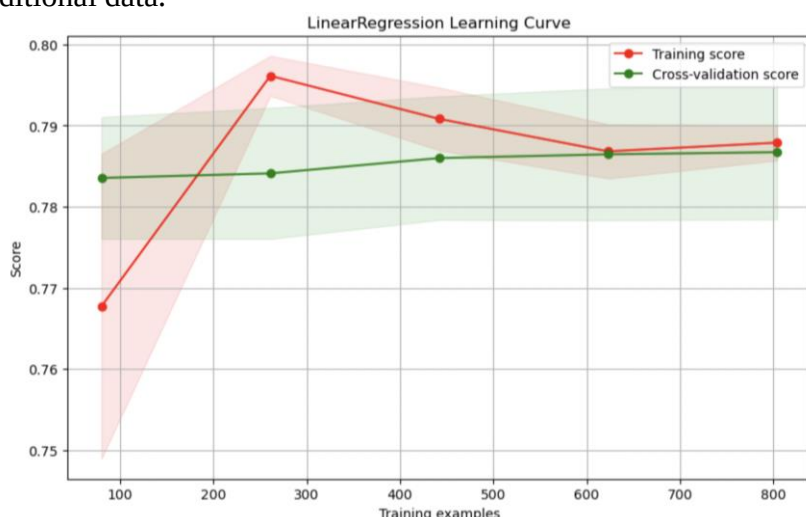


Fig. 4 Linear Regression Learning Curve (Photo/Picture credit: Original).

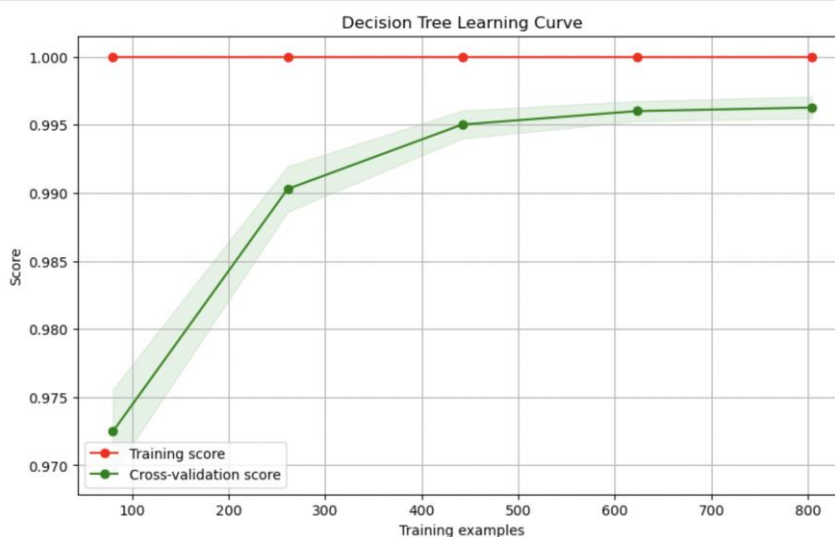


Fig. 5 Decision Tree Learning Curve (Photo/Picture credit: Original).

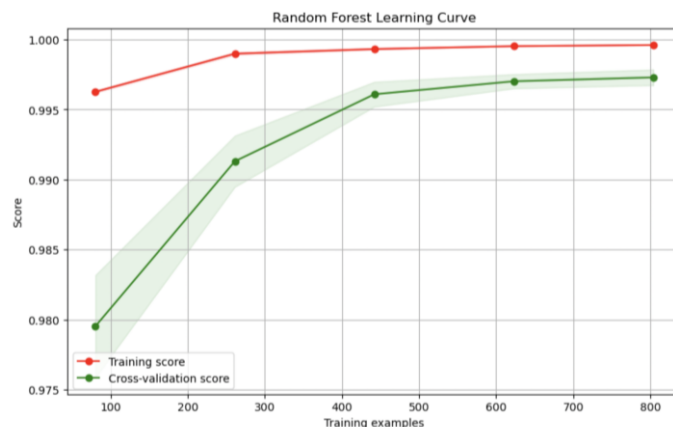


Fig. 6 Random Forest Learning Curve (Photo/Picture credit: Original).

4. Discussion

In the research, the researcher thoroughly evaluated the performance of several machine learning models in predicting stock prices. To measure the accuracy of these models, the researcher used the Mean Squared Error (MSE) as the primary metric. A lower MSE value indicates higher prediction accuracy, which means a minor difference between the predicted and actual stock prices.

After analyzing the data, the study found that the Random Forest Regressor model was highly effective. It consistently generated lower MSE values than the other models, which indicated its superior accuracy in forecasting stock prices. The researcher plotted the model's predictions against actual stock prices to understand its performance better, allowing for a direct visual comparison. The graphs supported the numerical results, and the Random Forest Regressor showed a stronger correlation to the actual stock price patterns.

The superior performance of the Random Forest Regressor is attributable to its capability of capturing complex, nonlinear patterns in stock market data, as emphasized in Kittipob Saetia's "Stock Movement Prediction Using Machine Learning Based on Technical Indicators and Google Trend Searches in Thailand" study [10]. Compared to a single Decision Tree model, the Random Forest Regressor has demonstrated high accuracy and low variance in research investigations. This suggests that the Random Forest model can make reliable predictions for the stock market by generalizing the training data to unseen data. This is particularly important because of the risk of overfitting and high variance in single decision tree models.

Additionally, the results support the statement made by Hera Antonopoulou in "Utilizing Machine Learning to Reassess the Predictability of Bank Stocks" regarding the effectiveness of Random Forests [11]. Random Forests, a collection of decision trees, naturally minimize the high variance arising from individual trees. Random Forests are a type of decision tree algorithm that can minimize overfitting by combining the predictions of multiple trees. This is especially useful in the stock market, where it's essential to avoid making extreme predictions based on abnormal data, mainly when only one data used to predict could cause a wrong prediction. By reducing these risks, Random Forests produce more consistent and dependable predictions.

The study has shown that advanced machine learning techniques like Random Forests can improve stock price predictions. The model can capture complex patterns and generalize them from training data, making them valuable tools in the financial market. However, the computational demands of the model must be carefully considered to strike a balance between prediction accuracy and practical feasibility.

5. Conclusion

The study has shown that the models used in machine learning are effective in predicting stock prices. Furthermore, the research has done a comparative analysis between these models. It demonstrated that machine learning techniques are effective in volatile markets like the stock market.

The literature review in the research has shown that machine learning techniques have a remarkable ability to predict market behavior. This trend highlights the current state of financial analysis and the increasing importance of data-driven algorithmic approaches.

The various evaluation metrics used in the research, such as mean square error, mean absolute error, and learning regression, have helped analyze the performance of each model. It revealed the difference between each model that was used in stock marketing. According to the analysis, the study has provided insights into the strengths and limitations of each model, which helped the researcher and financial worker choose a better-fit model for stock prediction.

The research has shown that machine learning significantly impacts the financial field, but it still needs to be continually refined to enhance predictive accuracy. This requires exploring more algorithms, integrating more data sets such as social media sentiment and economic indicators, and more advanced technologies.

The study has identified that Machine learning models are very effective at predicting stock prices. However, these models must evolve to keep up with changing financial markets. The ultimate goal remains to develop accurate, reliable, interpretable, and efficient models that meet the nuanced needs of contemporary financial analysis.

References

- [1] Song, Y., Lee, J. Importance of event binary features in stock price prediction. *Applied Sciences*, 2020, 10(5), 1597.
- [2] Ampomah, E. K., Nyame, G., Qin, Z., Addo, P. C., Gyamfi, E. O., Gyan, M. Stock market prediction with gaussian naïve bayes machine learning algorithm. *Informatica (Ljubljana)*, 2020, 45(2), 243-256.
- [3] Vijh, M., Chandola, D., Tikkiwal, V. A., Kumar, A. Stock closing price prediction using machine learning techniques. *Procedia Computer Science*, 2020, 167, 599-606.
- [4] Dahal, K. R., Pokhrel, N. R., Gaire, S., Mahatara, S., Joshi, R. P., Gupta, A., Banjade, H. R., Joshi, J. A comparative study on effect of news sentiment on stock price prediction with deep learning architecture. *PloS One*, 2023, 18(4), e0284695-e0284695.
- [5] Thakkar, A., Chaudhari, K. Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions. *Information Fusion*, 2021, 65, 95-107.
- [6] Sheng, Y., Ma, D. Stock index Spot–Futures arbitrage prediction using machine learning models. *Entropy (Basel, Switzerland)*, 2022, 24(10), 1462.
- [7] Agrawal, S. C. Deep learning based non-linear regression for stock prediction. *IOP Conference Series. Materials Science and Engineering*, 2021, 1116(1), 12189.
- [8] Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 2020, 297(3), 1178-1192.
- [9] Gao, Y., Kruger, U., Intes, X., Schwaitzberg, S., De, S. A machine learning approach to predict surgical learning curves. *Surgery*, 2020, 167(2), 321-327.
- [10] Saetia, K., Yokrattanasak, J. Stock movement prediction using machine learning based on technical indicators and google trend searches in thailand. *International Journal of Financial Studies*, 2023, 11(1), 5.
- [11] Antonopoulou, H., Theodorakopoulos, L., Halkiopoulos, C., Mamalouglou, V. Utilizing machine learning to reassess the predictability of bank stocks. *Emerging Science Journal*, 2023, 7(3), 724-732.