RESEARCH ARTICLE OPEN ACCESS

Predicting Stock Prices Using Machine Learning

Ahmet YILDIRIM[1], Ali GÜNEŞ[2]
[1]*(Department of Computer Engineering, Istanbul Aydın University, Turkey Email: ahmetyildirim5@stu.aydin.edu.tr)*
[2]*(Department of Computer Engineering, Istanbul Aydın University, Turkey Email: aligunes@aydin.edu.tr)*

ABSTRACT
Predicting the closing prices of stock instruments, which have become one of the prominent investment tools in our era, has become an important topic. Investors tend to use their investments to gain profit with minimal risk. Accordingly, predicting the closing prices of stocks based on future closing values is crucial. Two different methods are used for these decisions: fundamental analysis and technical analysis. With the rapidly developing software and hardware technology, the use of statistical methods in technical analysis is rapidly increasing. In this study, the closing values of APPLE, GOOGLE, MICROSOFT, AMAZON, and NVIDIA stocks between 2010 and 2019 were analyzed using Arima, XGBooster, Prophet algorithms, and LSTM (Long Short-Term Memory Networks).
Keywords: Machine Learning, Stock Prediction, Time Series, Artificial Neural Networks, Technical Analysis

---

Date of Submission: 02-07-2024 Date of acceptance: 12-07-2024

I. List of Abbreviations

1. **AI**: Artificial Intelligence
2. **ANN**: Artificial Neural Network
3. **ARIMA**: Autoregressive Integrated Moving Average
4. **GRU**: Gated Recurrent Unit
5. **LSTM**: Long Short-Term Memory
6. **MSE**: Mean Squared Error
7. **RMSE**: Root Mean Square Error
8. **MAE**: Mean Absolute Error
9. **OHLC**: Open, High, Low, Close
10. **BIST100**: Borsa Istanbul 100 Index
11. **AAPL**: Apple Inc. stock symbol
12. **GOOGL**: Alphabet Inc. (Google) stock symbol
13. **MSFT**: Microsoft Corporation stock symbol
14. **AMZN**: Amazon.com Inc. stock symbol
15. **NVDA**: NVIDIA Corporation stock symbol

II. INTRODUCTION

The stock exchange, one of the basic investment tools, is where shares belonging to companies, approved by the board of directors, are offered to the common market for buying and selling, allowing investors to trade. The valuable securities acquired by investors to become shareholders in a company are called stocks. Stocks provide investors with a certain share of ownership in the company's capital and the right to participate in the company's profits [1]. Stock investments can be short-term (less than a year), medium-term (between one and five years), or long-term (more than five years). Stocks traded on the stock exchange are grouped according to certain features into indices representing specific market segments or sectors, reflecting the performance of these stocks. Investors can buy and sell these indices just like stocks [2]. Indices offer investors the opportunity to diversify their portfolios and reduce risks compared to individual stocks [2]. However, investors use various methods and models to further reduce risks. These methods include portfolio diversification, asset allocation, and risk management strategies. For example, Modern Portfolio Theory (MPT) aims to determine the optimal portfolio distribution by analyzing the risk and return characteristics of different asset classes [2]. Additionally, derivative products and hedge strategies help investors protect against market fluctuations. These approaches allow investors to minimize risk and achieve more balanced and predictable returns [2].

Fundamental and Technical Analysis: Fundamental analysis involves examining the sector of the stock, the company's financial condition, future financial expectations, the economic program, and the country's economic situation to decide whether to invest in the stock. Technical analysis involves analyzing the stock's past price movements and financial results to make investment decisions. Developments in the hardware sector, in line with the advancing software industry, have allowed the use of machine learning alongside basic statistical methods. These methods stand out for their ability to adapt to complex and dynamic market conditions, offering more successful results compared to traditional forecasting approaches [3]. Machine learning is the process of creating software models capable of making predictions such as classification, clustering, or regression through mathematical and statistical operations based on past data. This process aims to more accurately predict future probabilities by processing and analyzing data [4]. Various approaches are used in machine learning for data analysis and processing. Classification focuses on determining which class a data point belongs to, while clustering aims to find the nearest cluster based on the data's features. Regression seeks to identify the trend and direction of data movements, predicting possible future values. Time series are data sequences that track changes in variables over time. These series can be at different frequencies like year, month, or day and are obtained by regularly recording observed events over time. Time series analysis examines the trends, seasonality, and other dynamics in the data over time, enabling more accurate predictions [5]. Time series analysis includes the logic of predicting future values using past data. This method examines the behavior of data over time, predicting future trends, seasonal changes, and other dynamics. During the analysis process, previous period data are carefully examined, and patterns and relationships obtained from these data are used to predict future values [6]. Stock closing value prediction is a regression analysis within time series analysis.

III. Literature Review

Literatür Taraması

Machine learning is a rapidly evolving discipline that aims to extract meaningful information from large data sets and make predictions based on this information. This study reviews important research conducted in the field of machine learning, evaluating their methodologies and results in detail. Machine learning allows computer systems to learn from data through specific algorithms and make predictions about future data based on this learning process. Various learning types, such as supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning, form the building blocks of this process. The reviewed research demonstrates the effectiveness of machine learning techniques and identifies the conditions under which they yield the best results.

Relevant Studies

- Salim DEMİRAY (2022): "Stock Closing Prediction Using Machine Learning Algorithms" focused on predicting Microsoft stock closing values using machine learning methods. Polynomial Regression, ARIMA, XGBoost, LSTM, and Facebook Prophet algorithms were used. The data set included daily closing values of Microsoft stocks from 2014-2019, obtained from Yahoo Finance. The study applied these methods using Python and various software packages, predicting stock closing prices for 5, 10, and 20-day periods and comparing their performances. Polynomial Regression provided the best prediction results, ARIMA showed generally good results despite increasing errors with longer prediction periods, XGBoost performed poorly, LSTM was successful in 20-day predictions despite higher errors in short-term predictions, and Facebook Prophet provided the worst predictions [14].
- İbrahim TUNA (2023): "Predicting Stock Data Using Machine Learning Methods" focused on predicting stock data using deep learning methods. Raw stock data obtained from Yahoo Finance were cleaned for "null" rows on non-trading days. Two models were developed: one using OHLC (Open, High, Low, Close) values and the other using only the closing value. Each stock index was tested separately, predicting closing values. The study measured test accuracy rates for three-day predictions using OHLC data for BIST100, Nikkei 225, and Nasdaq indices, with test accuracies of 0.520, 0.495, and 0.583, respectively. The second model, using closing and date columns, showed test accuracy rates ranging from 91% to 96% for BIST100, 78% to 89% for Nikkei 225, and 96% for Nasdaq [15].
- Sami KARACAN (2021): "Using Machine Learning and Artificial Intelligence for Stock Price Prediction" examined predicting stock prices using machine learning and artificial intelligence. The study showed that AI-based models could capture complex patterns related to future returns using various input variables. However, most studies were optimized for a limited number of securities or indices in a single country, and it remained unclear whether signals profitable in one country would be profitable in others. The complexity of predicting stock prices is due to numerous economic, political, and psychological factors. Investors aim to maximize returns by buying and selling stocks at the right time, but stock data often have a complex and non-linear structure. The use of AI in financial markets is a rapidly developing field. The increasing economic globalization and advances in information technology make predicting stock prices more challenging. This study emphasized how AI and machine learning studies could assist investors. The use of AI in portfolio management, robo-advisory, and finance will continue to grow in importance, requiring support from new studies [16].

- Sharma Prashant (2024): The study examined predicting Microsoft's stock prices using LSTM and Random Forest algorithms. Using two different machine learning methods, the study aimed to predict the realized values of stock prices. LSTM, with its capacity to consider past information in time series data to predict future values, and Random Forest, with its strong tree-based structure, aimed to improve data understanding and prediction accuracy. The study was significant for investors using technical analysis, contributing to more accurate predictions of stock price movements [17].
- Biswal Avijeet (2024): This study detailed the use of machine learning to predict stock prices, using historical data from 2012 to 2017 to predict 2017 stock prices. The LSTM model was chosen for its ability to learn long-term dependencies in time series data. The study provided a detailed guide on using the LSTM model, showing a strong similarity between predicted stock prices and actual prices, indicating that the model could be a reliable tool for investors and researchers. The study significantly contributed to the literature by demonstrating the applicability and effectiveness of machine learning methods in financial markets [17].
- Shahi Tej Bahadur et al. (2020): The study compared the performance of LSTM and GRU models in predicting stock prices and examined the impact of financial news sentiment on predictions. Both models performed similarly using only stock market features. However, the performance of both models improved significantly when financial news sentiment was added. LSTM and GRU models offered better predictions than individual models, achieving the expected target in the study [19].

IV. Objective

This study aimed to determine the performance and accuracy of time series algorithms to provide the technical analysis needed by long-term investors. The period between 2010 and 2019 was chosen as it was a time when significant economic fluctuations and market conditions did not significantly affect these companies' stock prices. This provided a suitable ground for evaluating the performance of the algorithms more consistently and reliably. Furthermore, considering the tendency of fundamental analysis factors to significantly impact stock prices during this period, these factors were analyzed and evaluated periodically. Consequently, the selection of leading technology companies such as APPLE, GOOGLE, MICROSOFT, AMAZON, and NVIDIA was made based on their solid financial structures, innovative products, and long-term growth potentials. The analysis of these companies' stocks provided a suitable sample to determine the performance and accuracy of time series algorithms.

V. Method

The closing prices of APPLE, GOOGLE, MICROSOFT, AMAZON, and NVIDIA stocks between 2010 and 2019 were obtained from Yahoo Finance and used in this study. The closing values read from the data set were used to train time series machine learning algorithms using values between 2010 and 2019, predicting the 2019 values and comparing them with the actual values in 2019 to find the most suitable algorithm. The ARIMA, XGBooster, LSTM (Long Short-Term Memory Networks), and Prophet algorithms used values between 2010 and 2018 for training and compared the 2019 values with the actual values. This study used daily closing values of Apple, Google, Microsoft, Amazon, and NVIDIA stocks obtained from Yahoo Finance. The machine learning algorithms trained with data from 2010 to 2018 predicted the actual values in 2019, and the values produced by the algorithms were compared with the actual values in 2019, calculating the deviation. The study aimed to find the closest results to the actual stock closing prices. Anaconda navigator-3, Spyder 5.4.3 IDE, and Python 3.11.7 package were used for the study, applying various software packages to examine prediction results using four different methods.

ARIMA Method The Autoregressive Integrated Moving Average (ARIMA) is an effective method widely used in time series analysis. This model is applied after making the series stationary through differencing or multiplication. The ARIMA(pdq) model relies on three fundamental parameters: the 'd' parameter indicates the number of times the series needs to be differenced to become stationary, the 'p' parameter indicates the degree of the autoregressive component, and the 'q' parameter indicates the degree of the moving average component [7]. The autoregressive component (p) models the current value as a function of previous values, representing the dependency of the next value in the series on previous values.

XGBooster Method XGBoost, developed by Tianqi Chen and Carlos Guestrin, is an improved decision tree algorithm in terms of speed and performance [8]. The algorithm can work quickly with its prediction power, missing data management, and overfitting prevention. Gradient Boosted weak leaves, a fundamental part of the algorithm, aim to transform step-by-step into stronger leaves, producing better results [8]. Decision trees, represented by a graph, help in finding solutions by grouping data based on conditions. Grouping is called a tree because it starts from a root and branches out into sub-roots and leaves.

LSTM (Long Short-Term Memory) Method Traditional artificial neural networks (ANN) lacked the ability to behave according to the previous state in time series analysis. To address this deficiency, the Long Short-Term Memory (LSTM) method was developed [9]. ANNs tend to forget past information and focus only on the current state because they do not adequately consider temporal dependencies in the data they are trained on. LSTMs, with their self-feeding structures like a chain, can retain information from past data, learn long-term dependencies, and use this information effectively in future predictions.

Prophet Method Prophet, developed by Facebook, is a library published to automatically predict time series data considering special periods such as holidays [10]. Prophet takes two variables, time and value, and is robust against missing and floating data. It supports trend, seasonality, and holidays in time series prediction [11].

Error Measurement Method To evaluate the measurements, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE) methods were used to measure the variation between the predicted value and the actual value. MAE is the value obtained by summing the differences between the actual and predicted values. MSE is the value obtained by summing the squares of the differences between the actual and predicted values. RMSE is the value obtained by taking the square root of the sum of the squares of the differences between the actual and predicted values [12].

Research Results Charts showing the closing prices of Apple, Google, Microsoft, Amazon, and Nvidia stocks were analyzed. Daily error results for ARIMA, XGBooster, LSTM (Long Short-Term Memory Networks), and Prophet methods are shown in Table 1. As seen in Chart 2, LSTM provided the lowest error value.

VI. Conclusion

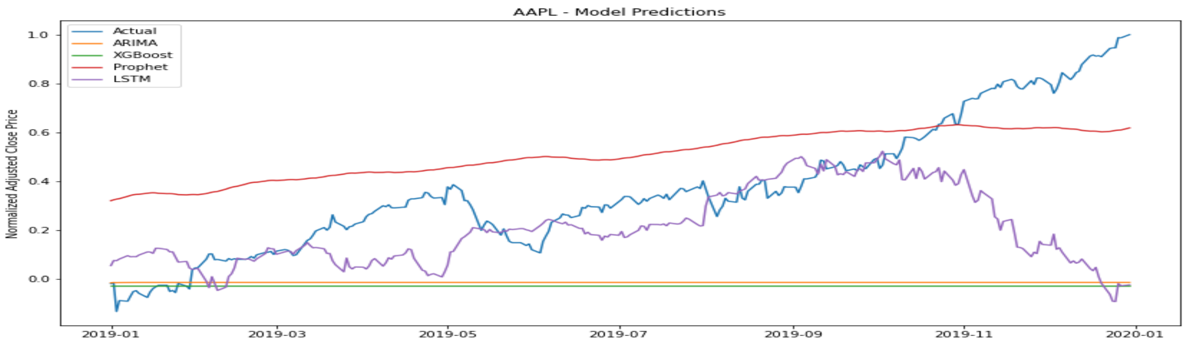


Chart -1 Apple Stock Closing Chart

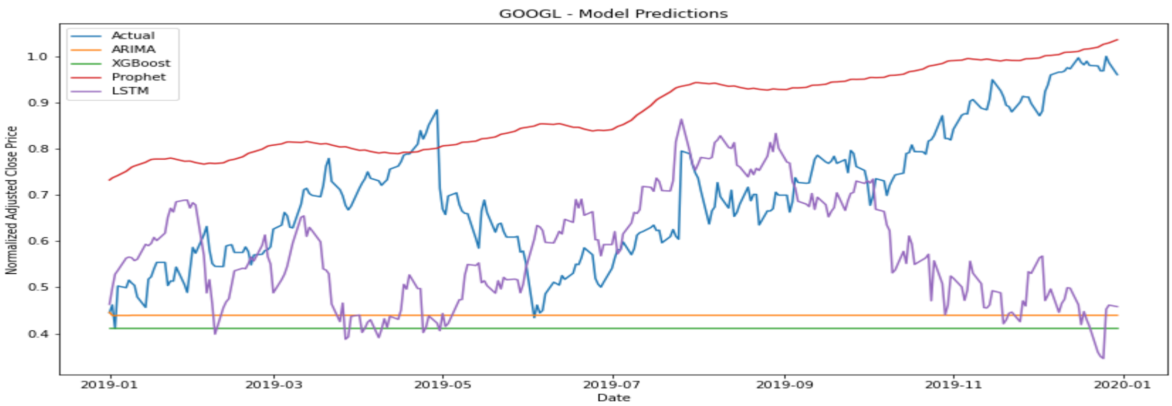


Chart -2 GOOGLE Stock Closing Chart

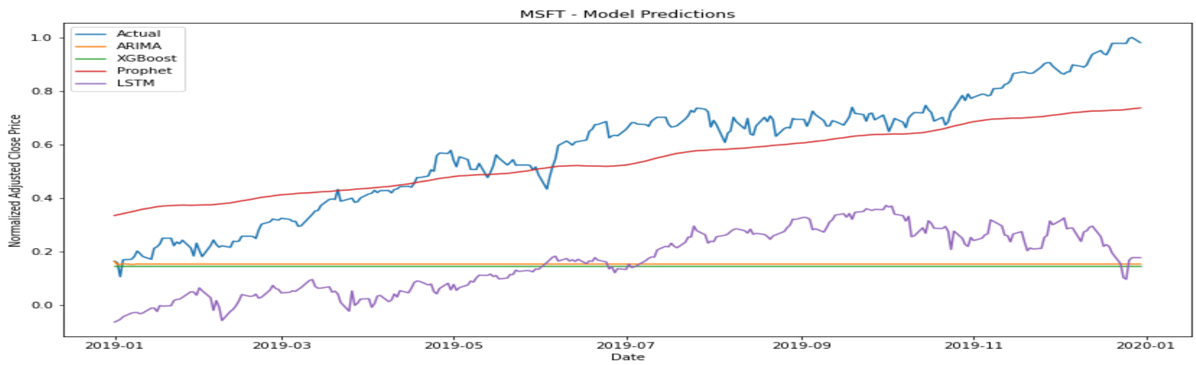


Chart -3 MICROSOFT Stock Closing Chart

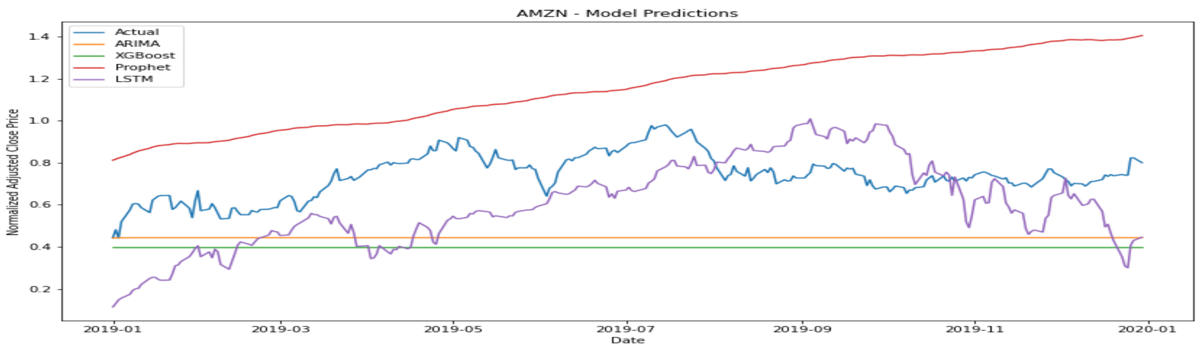


Chart -4 AMAZON Stock Closing Chart

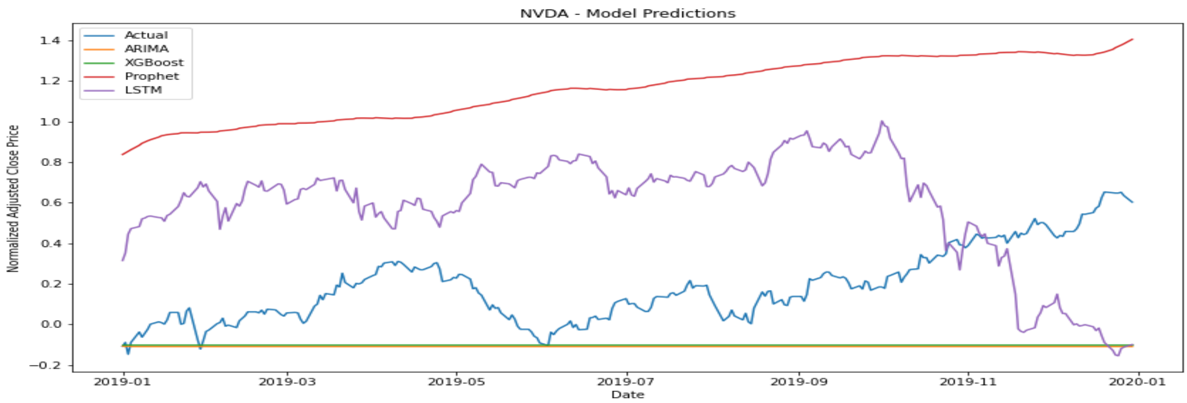


Chart -5 NVIDIA Stock Closing Chart

Algorithm	MSE	RMSE	MAE
AAPL ARIMA	0.211442	0.459829	0.383491
AAPL XGBoost	0.222753	0.471967	0.396123
AAPL Prophet	0.055112	0.234759	0.214185
AAPL LSTM	0.099618	0.315623	0.200072
AAPL GRU	0.102561	0.320251	0.204335
GOOGL ARIMA	0.086772	0.294571	0.259658
GOOGL XGBoost	0.102222	0.319722	0.287659
GOOGL Prophet	0.042171	0.205355	0.183390
GOOGL LSTM	0.059079	0.243062	0.188042
GOOGL GRU	0.061320	0.247629	0.190598

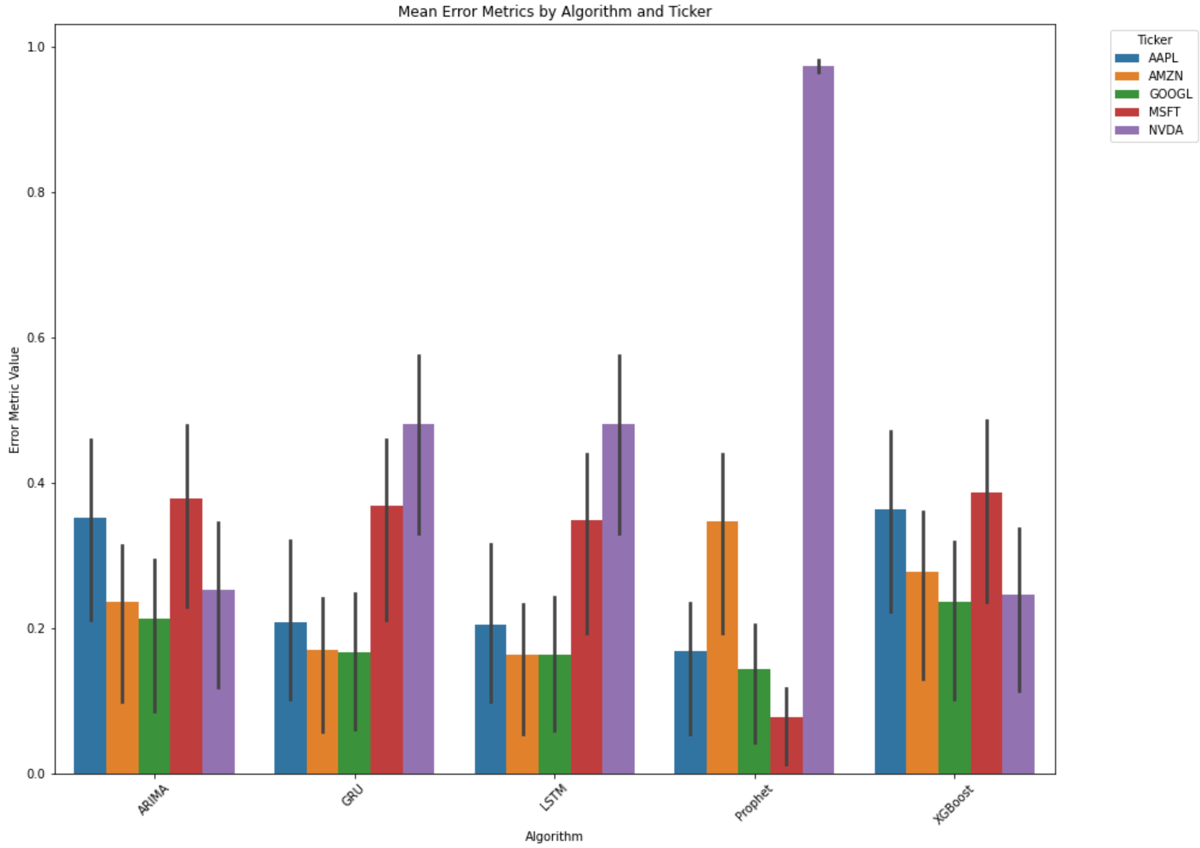
MSFT ARIMA	0.230019	0.479603	0.427178
MSFT XGBoost	0.236902	0.486726	0.435134
MSFT Prophet	0.013901	0.117904	0.101126
MSFT LSTM	0.192907	0.439211	0.415904
MSFT GRU	0.211009	0.459357	0.436499
AMZN ARIMA	0.098965	0.314588	0.296188
AMZN XGBoost	0.129861	0.360362	0.344399
AMZN Prophet	0.193607	0.440008	0.407884
AMZN LSTM	0.054403	0.233245	0.205195
AMZN GRU	0.058132	0.241106	0.211814
NVDA ARIMA	0.119010	0.344978	0.294960
NVDA XGBoost	0.114104	0.337793	0.286715
NVDA Prophet	0.966546	0.983131	0.973172
NVDA LSTM	0.331551	0.575804	0.538322
NVDA GRU	0.330873	0.575216	0.537865

Table 1: Prediction Error Values

MSE: Mean Squared Error

RMSE: Root Mean Square Error

MAE: Mean Absolute Error



Graph 2: Forecast Error Graph

As seen from the error chart, comparing the average error metrics of various algorithms (ARIMA, GRU, LSTM, Prophet, XGBoost) and stock symbols (AAPL, AMZN, GOOG, MSFT, NVDA) reveals that each bar's height represents the average error value for a specific algorithm and stock, while error bars represent the standard deviation.

The ARIMA algorithm generally showed moderate performance in terms of error metrics compared to other algorithms. Relatively lower error metrics were observed for Amazon (AMZN) and Google (GOOG) stocks.

The GRU (Gated Recurrent Unit) algorithm generally performed well, particularly showing low error values for AMZN and GOOG stocks.

The LSTM (Long Short-Term Memory) algorithm provided the best results for AMZN and GOOG stocks but showed higher error metrics for MSFT and NVDA stocks, indicating less effectiveness on these stocks.

The Prophet algorithm showed very high error metrics for MSFT stock, indicating that it was unsuitable for this stock. It showed average performance for other stocks.

The XGBoost algorithm generally performed well with low error metrics but showed higher error metrics for MSFT stock compared to other algorithms.

These results demonstrate that different algorithms exhibit varying performances on different stocks. The LSTM model performed particularly well on AMZN and GOOG stocks but was less successful on MSFT and NVDA stocks. The Prophet algorithm showed high error metrics for MSFT stock. These findings suggest that algorithms should be selected based on stock characteristics, and the most suitable model should be determined for each stock.

In light of this analysis, it was found appropriate to use the LSTM model for stocks where it showed superior performance (AMZN and GOOG), but different algorithms should be used for other stocks. Future studies should include different variables and data to improve model performance and achieve more reliable predictions.

References

[1]. YILDIRIM Hakan and Murat KESEBİR. "Causal Relationships Between Stock Indices in Countries with Different Sizes of Economies and Financial Markets." Fırat University Journal of Social Sciences 29.1 (2019): 249-259.

[2]. Filiz E. KARABOGA H. A. & AKOGUL S. (2017). Using Machine Learning Methods and Artificial Neural Networks for Classifying BIST-50 Index Change Values. Journal of the Cukurova University Institute of Social Sciences 26(1)

[3]. Egeli B. (2003). Stock market prediction using artificial neural networks. Decision Support Systems 22 171-185.

[4]. Harwich E. & Laycock K. (2018). Thinking on its own: AI in the NHS. Reform Research Trust.

[5]. Rounaghi M. M. & Zadeh F. N. (2016). Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: monthly and yearly forecasting of time series stock returns using ARMA model. Physica A: Statistical Mechanics and its Applications 456 10-21

[6]. Akgül I. (2003). Traditional Time Series Methods. Baskı Der Yayınları Istanbul.

[7]. Tseng F. M. Yu H. C. & Tzeng G. H. (2002). Combining neural network model with seasonal time series ARIMA model. Technological forecasting and social change 69(1) 71-87.

[8]. https://xgboost.readthedocs.io/en/latest/tutorials/model.html

[9]. http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[10]. https://facebook.github.io/prophet/docs/trend_changepoints.html#automatic-changepoint-detection-in-prophet

[11]. Taylor S. J. & Letham B. Forecasting at scale. PeerJ Preprints 5: e3190v2 (2017).

[12]. Willmott C. J. & Matsuura K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research 30(1) 79-82.

[13]. Franses P. H. (2016). A note on the mean absolute scaled error. International Journal of Forecasting 32(1) 20-22.

[14]. Demiray Salim et. al. International Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622 Vol. 11 Issue 1 (Series-III) January 2021 pp. 59-62

[15]. Tuna I. & KASAPBAŞI M. C. (2023). Predicting Stock Data Using Machine Learning Methods. 9th International Congress on Engineering and Technology Management Turkey

[16]. Karacan Sami. (2021). Using Machine Learning and Artificial Intelligence for Stock Price Prediction. Journal of International

Social Research. 14. 666-674. 10.17719/jisr.11335.
[17]. Sharma Prashant (2021 October 5). Machine learning for stock market prediction with step-by-step implementation. Analytics Vidhya
[18]. Biswal A. (2024 April 15). Stock market prediction using machine learning in 2024. Simplilearn. Retrieved July 1 2024 from https://www.simplilearn.com/tutorials/machine-learning-tutorial/stock-price-prediction-using-machine-learning
[19]. Shahi TB Shrestha A Neupane A Guo W. Stock Price Prediction Using Deep Learning: A Comparative Study. Mathematics. 2020; 8(9):1441. https://doi.org/10.3390/math8091441