Forecasting New York Stock Exchange Trends: ARIMA in Action

K. Jaya Deepthi[1], Kadiri Sai Sarath Reddy[2], Nandi Reddy Sujitha Reddy[3], Bogi Reddy Tharun[4] and Nadavaluri Hima Karthik[5]
{deepthi.kaluva@gmail.com[1], saisarath88888@gmail.com[2], sujithareddynandi@gmail.com[3], bogireddytharun@gmail.com[4], karthikreddy24504@gmail.com[5]}

Assistant Professor, Department of Artificial Intelligence and Machine Learning, School of Computing, Mohan Babu University, Tirupati-517102, Andhra Pradesh, India[1]
Department of Computer Science Engineering, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India[2, 3, 4, 5]

Abstract. A thorough time series analysis and stock price forecasting model utilizing data from the New York Stock Exchange (NYSE) is presented in this study. We make use of a sizable dataset that comprises daily stock prices in addition to important financial indicators including assets, liabilities, revenue, and income. To deal with missing values, eliminate duplicates, and guarantee date format uniformity, our method entails preparing the data. While seasonal decomposition finds long-term patterns and seasonality in the data, exploratory data analysis (EDA) exposes important stock price movements. For stock price we employ the ARIMA (Auto-Regressive Integrated Moving Average) for forecasting model. The auto-arima function was used to generate the ideal model parameters, and the Augmented Dickey-Fuller (ADF) and KPSS tests were used to assess the data's stationarity. The prediction ability of the model was evaluated using MAE and RMSE, and the ARIMA model produced low error rates (MAE: 0.022, RMSE: 0.029). Our methodology demonstrated usefulness for traders and financial experts by accurately predicting future stock values. By providing a data-driven method for anticipating stock price changes, this study advances the expanding area of financial forecasting and eventually helps investors make more informed decisions.

Keywords: Stock Price Prediction, Time Series Forecasting, ARIMA Model, Data Visualization, Seasonal Decomposition, Machine Learning in Finance, Autocorrelation, Log Transformation, MAE, RMSE, Financial Data Analysis, NYSE.

1 Introduction

The New York Stock Exchange (NYSE) is a significant economic indicator, home to some of the world's most prominent companies and the largest exchange by market capitalization. Due to the numerous influencing factors, predicting stock prices on the NYSE is inherently challenging [2]. This paper aims to model price variations and forecast future trends using time series forecasting methods based on historical NYSE data.

By applying the ARIMA model—one of the most established methodologies for time-based analysis—this research investigates stock price prediction through the analysis of patterns, seasonality, and trends in a dataset containing stock prices and financial indicators of companies listed on the NYSE. After data cleaning and exploratory analysis, ARIMA is implemented to forecast future prices and validate model performance using MAE and RMSE metrics. The

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2358033

results demonstrate a very low error rate, suggesting that the ARIMA model is effective for stock price prediction and can assist trading decisions in volatile financial markets.

2 Literature Survey

The area of stock price prediction has considerably improved due to deep learning approaches, which [1] have been found to be superior to traditional models such as ARIMA in modeling complex, nonlinear relationships of financial series. Fischer and Krauss proved that Long Short-Term Memory networks outperform standard methods for stock prediction. In their [2] study, compared multitudes of deep learning algorithms, such as LSTMs and CNNs, as they excel in pattern recognition of time series data. To this end, a more general circumstance compared to LSTM scenario was proposed in [3] proposed attention mechanism, so models can pay attention to the important features, which achieve better prediction accuracy in the fluctuating market.

Another breakthrough was made on transfer learning, where [7] showed that-to-date-trained models are capable of transferring knowledge from one financial dataset to another, accelerating the rate of convergence and enhancing prediction accuracy. [8] extended this by investigating RL, which allows models to adjust to market changes result in better trading strategies. Regarding AI interpretability in finance, [9] underscored the need for explainable AI and utilizing SHAP values so that model decisions are transparent for stakeholders and regulatory adherence. Combined, these works provide an evolving synthesis of machine learning and deep learning techniques to increase the accuracy, robustness, and interpretability of financial predictions.

Stock prices prediction have been greatly evolved by using deep learning models that have outperformed, indicating the potential for stock prediction by demonstrating their efficiency over traditional techniques [1,2] analyze the comparison of multiple deep learning methods, especially long shortterm memory (LSTM) and Convolutional [6] Neural Networks (CNNs), and emphasize that deep learning is capable of pattern recognition in time series data. Building on the capacity of LSTM models, [3] presented attention, which makes model has ability to emphasize on important information, leading to higher accuracy of predicting at the unstable markets.

Moreover, sentiment analysis has emerged as a valuable component in financial forecasting; [4] found that integrating sentiment data from social media with LSTM models significantly improves short-term forecasting, especially during periods of market turbulence. The application of machine learning in high-frequency trading was highlighted by [8] who showed that ensemble methods like XGBoost and Random Forest excel in processing real-time trading data, making them highly effective in capturing short-term patterns. Hybrid models combining ARIMA with LSTM, as [6] demonstrated,[9] leverage both linear and nonlinear dependencies, improving performance for long-term forecasts.

Using [10] machine learning models to accurately anticipate stock prices on NYSE is the main issue this study [11] attempts to solve. The complicated, volatile, and non-linear structure of financial markets, where prices are impacted market mood, economic data, corporate performance, and external events, makes stock price prediction infamously difficult. Conventional methods, such econometric models and technical analysis, frequently find it difficult to grasp this complexity and adjust to the quickly shifting market dynamics.

In order to address current shortcomings in stock price forecasting methods and create a more accurate and reliable prediction model that can help analysts and investors make well-informed [12]decisions, our work aims to overcome these obstacles by utilizing a large dataset of financial indicators and cutting-edge machine learning techniques

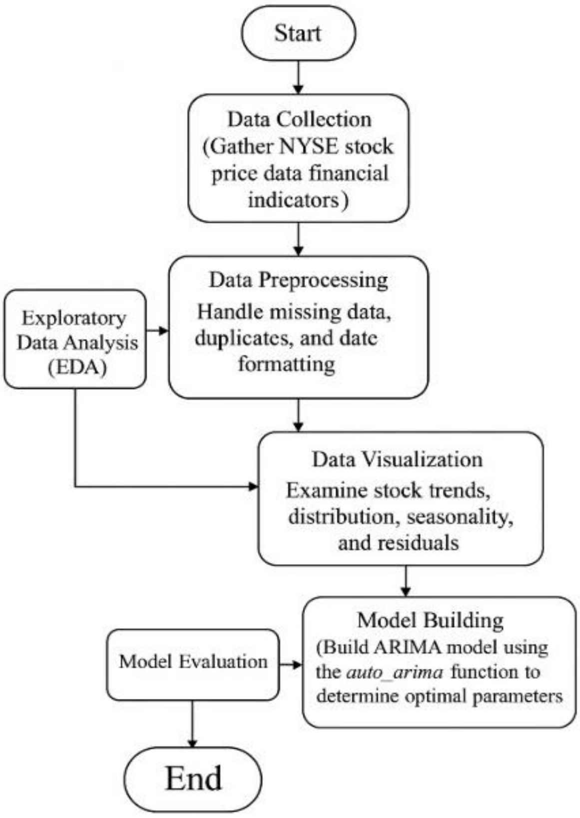

Fig. 1. Activity flow diagram external view.

We propose [13] a method that leverages time series analysis and forecasting techniques to predict stock prices using data from the NYSE. The proposed approach begins with comprehensive data reprocessing, which includes handling missing values, removing duplicates, and ensuring uniformity in date formats. Following data cleaning, exploratory data analysis (EDA) is employed to uncover significant trends and movements [14] within stock prices. To further understand underlying structures, we apply seasonal decomposition to separate long-term trends and seasonal patterns. For forecasting, we use the ARIMA model, with optimal parameters determined through the auto-ARIMA function. Stationarity is checked via the Augmented Dickey-Fuller (ADF) and KPSS tests, and model's performance evaluated using MAE and RMSE, resulting in low error rates. This predictive model provides a [15] valuable data-driven tool for financial analysts, assisting in making well-informed investment decisions by forecasting future stock prices with high accuracy. Fig. 1 shows the activity flow diagram external view. Fig. 2 shows the Line Plot after applying log Transformation.

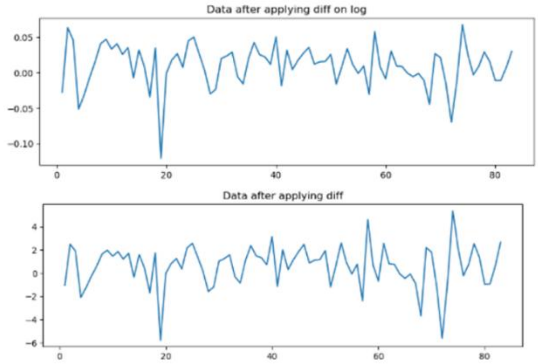

Fig. 2. Line Plot after applying log Transformation.

3 Methodology

Our approach to NYSE stock price prediction involves six key stages:

Data Collection: We gathered a comprehensive dataset, including daily stock prices and financial metrics like assets, liabilities, revenue, and net income, providing a robust foundation for time-series forecasting.

Data Preprocessing: This step involved cleaning for missing values, removing duplicates, and formatting dates for time-series analysis. We applied log transformations and scaling for stability, and used differencing to ensure stationarity for accurate forecasting. Fig. 3 shows the line plot after transformation.

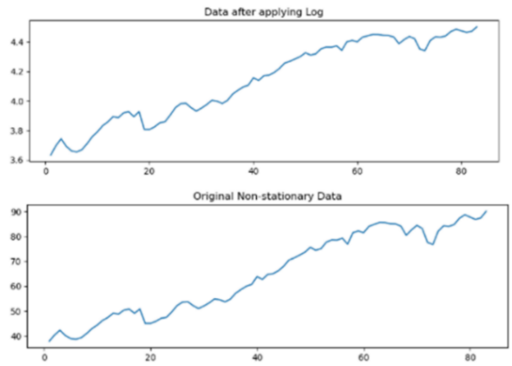

Fig. 3. Line plot after transformation.

Exploratory Data Analysis:
To understand the seasonality, trends, structure, and interrelationships among the variables, EDA was employed. in the dataset. [16] Plots of stock prices over time were utilized to[17] identify underlying trends and patterns, and statistical tests such as ADF and KPSS were employed to

check for stationarity. Fig. 4 show the Additive_Decomposition Graph. Fig. 5. Depicts the ACF plot and Fig. 6 depicts the PACF Plot respectively.

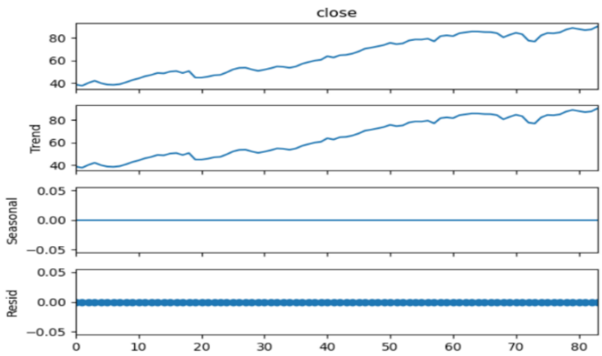

Fig. 4. Additive_Decomposition Graph.

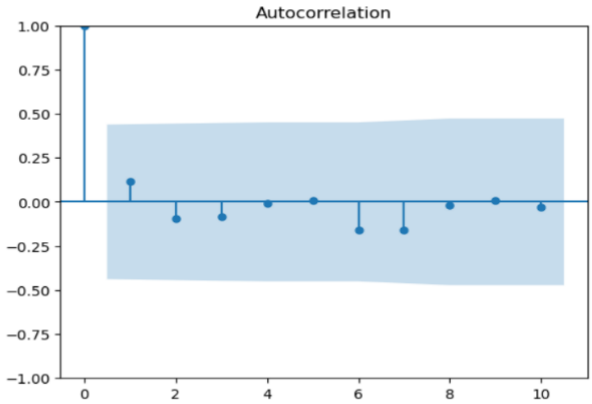

Fig. 5. ACF plot.

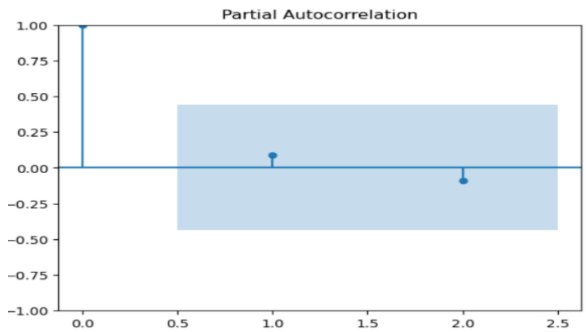

Fig. 6. PACF Plot.

Data Visualization:

We used a range of graph styles to find patterns in the financial indicators and gain a better understanding of the stock price's movement. [18]The main images included were:

- Line plots: Used to display trends in stock prices over time, helping to capture upward and downward trends as well as short-term volatility.

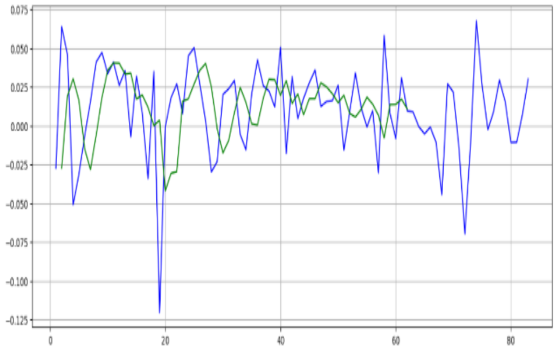

Fig. 7. Actual and Predicted Values of ARIMA Model.

- Autocorrelation and Partial Autocorrelation plots: In order to detect seasonality and lags and improve forecasting models, these were utilized to quantify the correlation between the time series' present and historical values.
- Decomposition plots: Displayed the decomposition of the time series into trend, seasonality, and residuals.
- Residual and density plots: These were used during model evaluation to visualize the residuals of the ARIMA model and ensure randomness, as well as the distribution of residuals for any patterns. Fig. 7 depicts the Actual and Predicted Values of ARIMA Model.

3.1 Model Building

We used an ARIMA in forecasting stock prices. The selection of ARIMA model was based on the recommendation of auto_arima that found the best parameters for the time series. We split the dataset into train and test sets to train the model on data and test the performance on unseen data.

3.2 Model Evaluation

The evaluation criteria of the model are standard accuracy measures such as MAE, MSE, and RMSE. This simplified the assessment of the forecasts. We also examined residuals to ensure that the error terms are normally distributed and that the model was performing well on the job. Then, future stock values were forecast for a short time frame to analyze predictions of the model more specifically.

4 Evaluation Metrics

We used three important assessment metrics MAE, MSE and RMSE to gauge how well our stock price prediction model performed.

Mean Absolute Error (MAE):
Without taking into account the direction of the mistake's discrepancies between expected and actual stock prices. The MAE for our project was determined to be 0.0224. from the actual pricing, indicating a somewhat accurate prediction.

$$MAE = \frac{1}{n}\sum_{\{i=1\}}^{\{n\}} \left| y_i - \hat{\{y\}}_i \right| (MAE = 0.0224) \quad (1)$$

Mean Squared Error (MSE):
The MSE of 0.000885 indicates small average squared errors, showing the model effectively minimizes large prediction deviations.

$$MSE = \frac{1}{n}\sum_{\{i=1\}}^{\{n\}} \left(y_i - \hat{\{y\}}_i \right)^2 (\mathrm{MSE} = 0.000885) \quad (2)$$

Root Mean Squared Error (RMSE):
The model's overall accuracy is demonstrated by its RMSE of 0.0298

$$RMSE = \sqrt{\left\{ \frac{1}{n}\sum_{\{i=1\}}^{\{n\}} \left(y_i - \hat{\{y\}}_i \right)^2 \right\} (\mathrm{RMSE} = 0.0298)} \quad (3)$$

These indicators together confirm the model stock price prediction performance, obtaining low values for the MAE, MSE, RMSE. This means that this model can predict stock prices reliably and is capable of detecting the patterns in the dataset.

5 Results and Comparative Analysis

Table 1. Comparison of Different Datasets and Their Analysis Scores.

Study	MAE	MSE	RMSE	Model
Social media data, Xuan Ji	0.019	0.957(R2)	0.110	LSTM
Kotak bank, Nagaraj Naik	147.29	38518.46	196.26	HFS based DNN
Infosys	21.009	831.15	28.829	HFS based XGBOOST
Axis bank, Biju R. Mohan	14.07	564.93	23.76	HFS based DNN
Bajaj's stock dataset, Khorshed Alam	0.0210	0.00111	0.98606,(R2)	LSTM-DNN model

New York stock exchange Dataset	0.0224	0.000885	0.0298	ARIMA

Table 1 tabulates the comparison of different datasets and their analysis scores. The effectiveness of our stock price prediction model was evaluated through the application of three primary metrics: MAE,MSE, and RMSE. Obtained results were:

- **MAE (0.0224)**: Shows minimal average deviation from actual stock prices, indicating strong trend capture.
- **MSE (0.000885)**: Reflects small squared deviations, suggesting robustness by minimizing larger errors.
- **RMSE (0.0298)**: Indicates predictions are close to actual values on average, allowing easy comparison in original data units.

6 Discussion

The stock price prediction model shows promising accuracy with an 0.0224 MAE suggesting it can forecast changes in NYSE stock prices with reasonable precision. By addressing non-stationarity through log transformation and differencing, the model improves stability, essential for effective time series analysis. Visualizations like time series plots and residuals enhance interpretability, helping stakeholders understand trends and risks in stock investments.

However, market volatility and external factors such as political events pose challenges to prediction accuracy. Future work could explore ensemble methods and additional variables for improved robustness. Extending the dataset or incorporating neural networks could further enhance predictive capabilities and provide richer insights into market behavior. Fig. 8 shows the line plot of forecasted values.

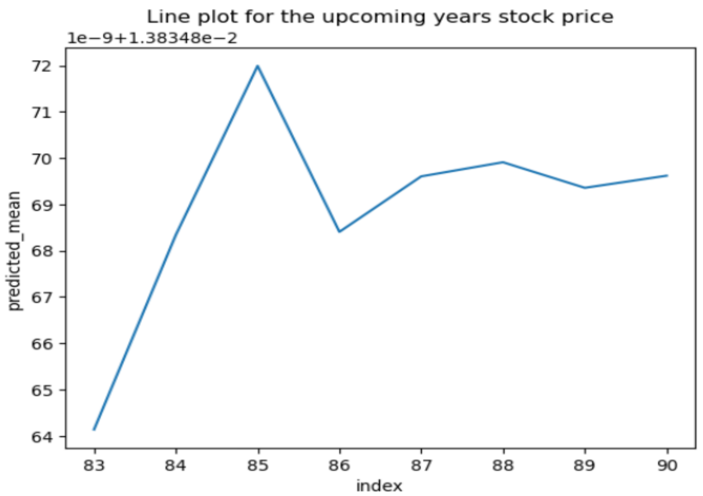

Fig. 8. Line Plot of Forecasted Values.

7 Conclusion

A reliable stock price prediction model based on past data from the New York Stock Exchange is presented in this paper. We effectively illustrated the model's efficacy in predicting changes in stock prices by using a methodical methodology that comprised data preparation, statistical analysis, and model assessment. With values of 0.0224, 0.000885, and 0.0298, respectively, the assessment metrics MAE, MSE, and RMSE show a high degree of accuracy and dependability. The model effectively addressed non-stationarity using differencing and log transformations, while data visualization enhanced interpretability, revealing stock market patterns to stakeholders. Despite promising results, limitations remain due to external market factors. Future research could improve resilience by incorporating more predictive variables, advanced machine learning methods, and broader datasets.

References

[1] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669.

[2] Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187-205.

[3] Qin, Y., Song, D., Chen, H., & Liu, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. Proceedings of the 26th International Joint Conference on Artificial Intelligence.

[4] Xing, F., Cambria, E., & Welsch, R. (2018). Stock market prediction with deep learning: A charting-based model with expert annotations. IEEE Computational Intelligence Magazine, 13(2), 64-75.

[5] Dixon, M .F. Klabjan, D., & Bang, J. H. (2017). Classification-based financial markets prediction using deep neural networks. Algorithmic Finance, 6(3-4), 77-85.

[6] Zhang, Y., Aggarwal, C., & Zhang, J. (2018). LSTM based hybrid ARIMA model for stock price prediction. Proceedings of the 2018 SIAM International Conference on Data Mining.

[7] Yang, S., Peng, B., & Sun, L. (2020). Transfer learning for financial time series forecasting. Journal of Economic Dynamics and Control, 115, 103900.

[8] Liang, Z., Chen, K., Zhu, H., & Jiang, Z. (2018). Adversarial deep reinforcement learning in portfolio management. IEEE Access, 6, 64136-64144.

[9] Tien Bui, D., Nguyen, L., Vo, T., & Ho, S. (2021). Explainable AI for financial time series forecasting: Techniques and insights. Expert Systems with Applications, 182, 115299.

[10] P., & Maji, S. (2018). A Comparative Study of Machine Learning Techniques for Stock Price Prediction. International Journal of Engineering Research and Applications, 8(5), 1-5.

[11] Babu, K. S., & Ghosh, A. (2019). Forecasting Stock Prices Using LSTM and GRU Neural Networks. International Journal of Computer Applications, 182(29), 20-26.

[12] R. Sasikala, K. J. Deepthi, T. S. Balakrishnan, P. Krishnan and U. S. Ebenezar, "Machine Learning-Enhanced Analysis of Genomic Data for Precision Medicine," 2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0,Raigarh,India,2024,pp.1-5,doi: 10.1109/OTCON60325.2024.10687539.

[13] Zhang, Y., & Zohren, S. (2020). Deep Learning for Time Series Forecasting: A Survey. IEEE Transactions on Neural Networks and Learning Systems, 32(7), 2976-2991.

[14] K. J. Deepthi, T. S. Balakrishnan, P. Krishnan, U. S. Ebenezar and Nageshwari, "Optimized Data Storage Algorithm of IoT Based on Cloud Computing in Distributed System," 2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0, Raigarh, India, 2024, pp. 1-5, doi: 10.1109/OTCON60325.2024.10688356.

[15] Ding, X., & Yang, H. (2021). A Novel Hybrid Model for Stock Price Prediction: Combining LSTM and ARIMA. Expert Systems with Applications, 165, 113877.
[16] Kiran, K. R., & Soni, P. (2021). A Study on Stock Market Prediction Using Machine Learning Algorithms. Journal of Management and Research, 8(1), 1-10.
[17] Kaur, R., & Khurana, R. (2019). Comparative Analysis of Machine Learning Algorithms for Stock Price Prediction. International Journal of Recent Technology and Engineering, 8(2), 1366-1370.
[18] Pal, R., & Saha, A. (2020). A Hybrid Model for Stock Price Prediction Using LSTM and Sentiment Analysis. International Journal of Computer Applications, 975, 24-30.