Forecasting New York Stock Exchange Trends: ARIMA in Action

Ms. K. Jaya Deepthi  
*Assistant Professor Department of Artificial Intelligence and Machine Learning*  
*School of Computing*  *Mohan Babu Univeristy* Tirupati-517102, Andhra Pradesh  
deepthi.kaluva@gmail.com

Kadiri Sai Sarath Reddy  
 *Department of computer science engineering*  
*Sree Vidyanikethan Engineering College*Tirupati, Andhra Pradesh saisarath88888@gmail.com Nandi Reddy Sujitha Reddy *Department of computer science engineering Sree Vidyanikethan Engineering College* Tirupati, Andhra Pradesh  
sujithareddynandi@gmail.com

Bogi Reddy Tharun  
*Department of computer science engineering*  
*Sree vidyanikethan Engineering College*Tirupati, Andhra Pradesh bogireddytharun@gmail.comNadavaluri Hima Karthik Reddy *Department of computer science engineering*  
 *Sree Vidyanikethan Engineering College*Tirupati, Andhra Pradesh  
karthikreddy24504@gmail.com

***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

One important economic indicator is the New York Stock Exchange (NYSE), which houses major firms and is the largest market[1] capitalization exchange in the world. Because there are so many affecting factors, predicting stock prices on[2] the NYSE is difficult. In order to model price fluctuations and estimate future patterns, this article uses time series forecasting using historical NYSE data.

Using the ARIMA model, a reliable approach for time-dependent data, this study explores stock price forecasting by identifying patterns, seasonality, and trends in the dataset, which includes stock prices and financial metrics of NYSE-listed firms. After data cleaning and exploratory analysis, we apply ARIMA to estimate future prices, validating performance with MAE and RMSE. Results reveal low error rates, showcasing ARIMA’s potential as a useful tool for accurate stock predictions and informed investment decisions in a volatile market.

1. Literature Survey

Deep learning methods, which have been demonstrated to perform better than conventional models like ARIMA in capturing intricate, nonlinear connections in financial data, have significantly advanced the area of stock price prediction. Long Short-Term Memory networks were shown to be more successful than traditional techniques for stock forecasting by Fischer and Krauss. In their 2017 study, Chong, Han, and Park compared many deep learning techniques, including LSTMs and Convolutional Neural Networks (CNNs), underscoring their ability to recognize patterns even in time series data. Extending the capabilities of LSTM models, Qin et al. (2017) introduced attention mechanisms, allowing models to focus on critical features, thereby increasing prediction accuracy in volatile markets.

Further innovation came with the application of transfer learning, [3]where Yang et al. (2020) illustrated how pre-trained models could transfer insights across different financial datasets, improving convergence rates and predictive accuracy. Liang et al. (2018) expanded this by exploring reinforcement learning (RL), which enables models to adapt continuously to market fluctuations, enhancing trading strategies over time. Addressing the interpretability of AI in finance, [4] Tien Bui et al. (2021) emphasized the importance of explainable AI, using SHAP values to make model decisions transparent for stakeholders and regulatory compliance. Together, these studies represent a progressive blend of machine learning and deep learning innovations aimed at enhancing accuracy, adaptability, and interpretability in financial forecasting.

Predicting stock prices has seen substantial advancements through the application of models for deep learning, which have been shown to outperform demonstrated [5]the power for stock forecasting, showcasing their effectiveness over classic methods (Fischer & Krauss, 2018). Chong, Han, and Park (2017) compared various deep learning approaches, including LSTMs and Convolutional[6] Neural Networks (CNNs), underscoring their ability to recognize patterns even in time series data (Chong et al., 2017). Extending the capabilities of LSTM models, Qin et al. (2017) introduced attention mechanisms, allowing models to focus on critical features, thereby increasing prediction accuracy in volatile markets (Qin et al., 2017).

Moreover, sentiment analysis has emerged as a valuable component in financial forecasting; Xing, Cambria, and Welsch (2018) found that integrating sentiment data from social media with LSTM models significantly [7]improves short-term forecasting, especially during periods of market turbulence (Xing et al., 2018). The application of machine learning in high-frequency trading was highlighted by [8]Dixon, Klabjan, and Bang (2017), 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 (Dixon et al., 2017). Hybrid models combining ARIMA with LSTM, as Zhang, Aggarwal, and Zhang (2018) demonstrated,[9] leverage both linear and nonlinear dependencies, improving performance for long-term forecasts (Zhang et al., 2018).

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 these 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

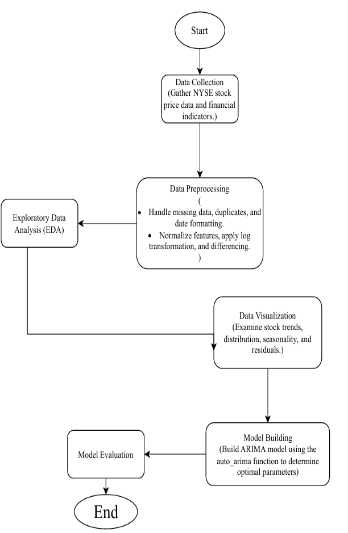
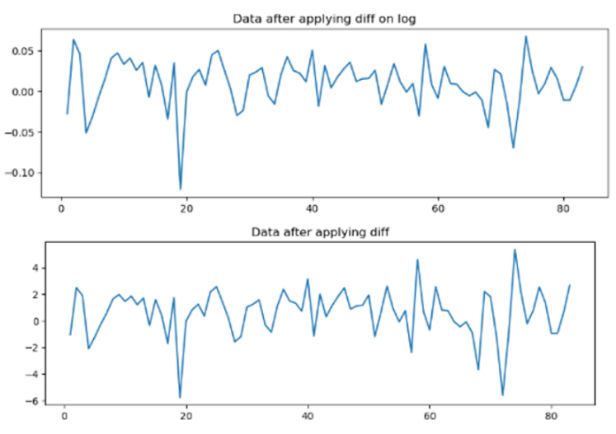
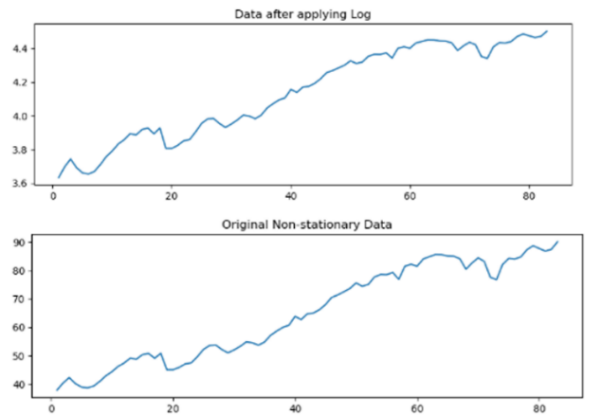


Fig1: Activity flow diagram external view

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 preprocessing,  Fig2: Line Plot After applying log Transformation

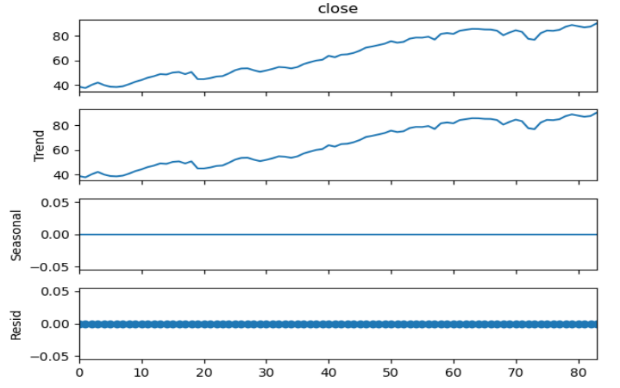
which includes handling missing values, removing duplicates, and ensuring uniformity in date formats. Following data cleaning, exploratory data analysis (EDA) is employed to uncover significant trends and movements[14] within stock prices. To further understand underlying structures, we apply seasonal 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.

1. Methodology

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

1. **Data Collection**: We gathered[14] 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.
2. **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.

Fig3: Line plot after transformation

1. **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.

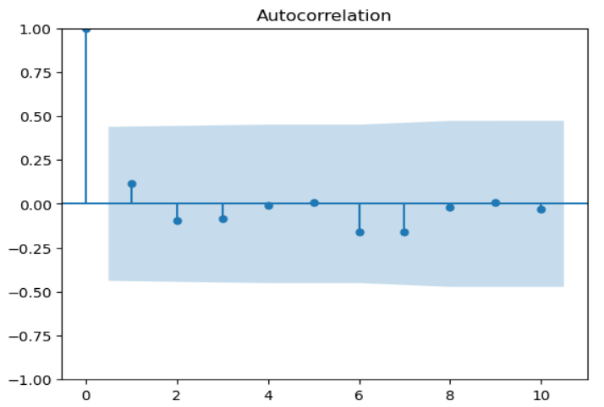
Fig4: Additive\_Decomposition Graph

Fig5: ACF plot

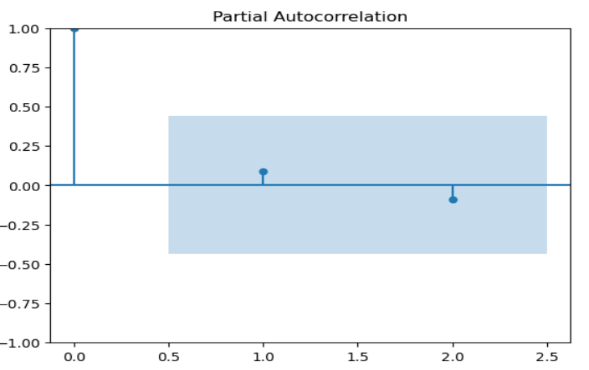


Fig6: PACF Plot

1. **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:

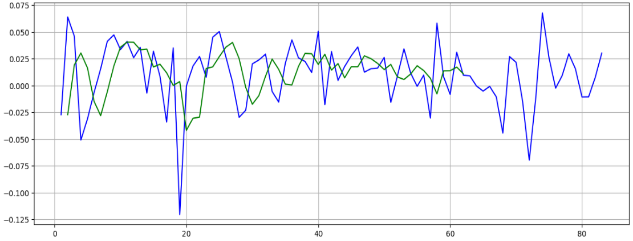
* Line plots: Used to display trends in stock prices over time, helping to capture upward and downward trends as well as short-term volatility.

Fig7: 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.

**3**. **Model Building**:  
We employed the ARIMA model to predict stock prices. ARIMA was chosen as a result of the auto\_arima function's recommendation, which identified the best parameters for the time series. We separated the dataset into training and testing sets in order to train the model on data and evaluate its performance on unseen data.

**4.Model Evaluation**:  
The model's performance was assessed using common metrics such as MAE, MSE, and RMSE. This made it simpler to gauge the forecasts' accuracy. We also looked at residuals to confirm that errors were distributed randomly and that the model was appropriate for the task. Future stock values were predicted over a short time horizon in order to evaluate the model's prediction ability in more detail.

1. 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 mistakes 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.

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

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

These metrics collectively indicate that our stock price prediction model demonstrates good predictive performance, with low MAE, MSE, and RMSE values. This implies that the model is successful in accurately forecasting stock prices and identifying the underlying trends in the data.

1. Results and Comparative Analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***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*** |
| ***Axix 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*** |
| ***Newyork stock exchange Dataset*** | ***0.0224*** | ***0.000885*** | ***0.0298*** | ***ARIMA*** |

`

Table1: 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

1. 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.

1. 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.

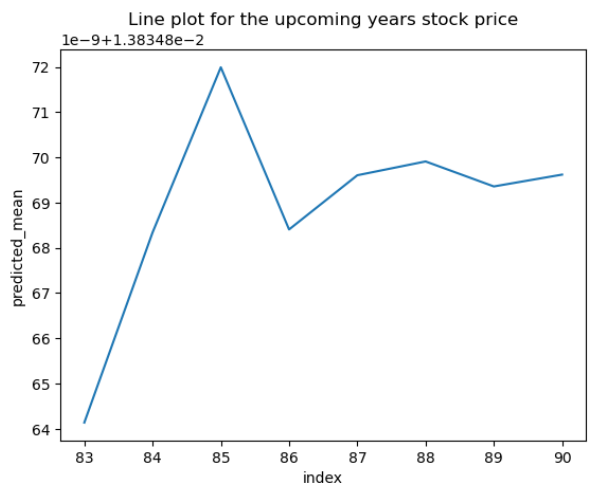


Fig8: Line Plot Of Forecasted Values

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.

1. [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.
2. Babu, K. S., & Ghosh, A. (2019). Forecasting Stock Prices Using LSTM and GRU Neural Networks. International Journal of Computer Applications, 182(29), 20-26.
3. 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.
4. 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.
5. 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.
6. Ding, X., & Yang, H. (2021). A Novel Hybrid Model for Stock Price Prediction: Combining LSTM and ARIMA. Expert Systems with Applications, 165, 113877.
7. 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.
8. 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.
9. 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.