Impact Factor (SJIF): 6.126 E-ISSN: 2583-1615, P-ISSN: 3049-3498

International Journal of Research in Academic World

Received: 24/May/2025 **IJRAW:** 2025; 4(7):46-49 **Accepted:** 02/July/2025

A Theoretical Approach to the Study of Stock Price Prediction-Comparative Analysis using ARIMA and Random Forest Model

***[1]Dr. Srinivas R, [2]Archana S and [3]Geethanjali**

*[1]Assistant Professor, Department of Management, Acharya Institute of Management and Science, Bengaluru, Karnataka, India.

[2]Assistant Professor, Department of Management and Commerce, S-Vyasa Advanced School of Studies (Deemed to be University), Bengaluru, Karnataka, India.

[3]Lecturer, Department of Computer Science, Yenepoya (Deemed to be University), Bengaluru, Karnataka, India.

Abstract

Stock price forecasting means estimating or forecasting the stock price of the future. Recently in the modern era, stock price forecasting is used for proper financial planning, efficient decision making, risk management and better investment planning. This study compares the performance of a machine learning algorithm (Random Forest) and a time series algorithm (ARIMA) in predicting stock prices using historical data from Apple Inc. From various financial sources, the data and information's are collected that majorly includes daily stock prices and features such as Open, High, Low, Close, Volume, and Adjusted Close. The study focuses on the Close prices as the target variable. Data pre-processing involves handling missing values and creating lag features (lag1, lag2, lag3) to enhance the predictive capability of the models. The entire data range are segregated into 80% training data and 20% testing data for evaluation. The ARIMA model, with parameters (p=5, d=1, q=0), is fitted to the training data and used to forecast stock prices for the test data. The Random Forest Regressor is trained on the lag features and used to predict stock prices. Both models are evaluated using Mean Squared Error (MSE) and Mean Absolute Error (MAE). The results indicate that the Random Forest model outperforms ARIMA in terms of both MSE and MAE, suggesting its better ability to capture complex relationships in the data. However, ARIMA remains valuable for scenarios requiring explainable and interpretable models. The study concludes that the choice between ARIMA and Random Forest should depend on specific requirements such as interpretability, accuracy, and computational cost. Future work includes combining both models and exploring advanced techniques like LSTM, GRU, and transformer networks for improved stock price forecasting.

Keywords: Stock price forecasting, ARIMA, MSE (Mean squared Error), MAE (Mean Absolute Error) random forest, machine learning.

Introduction

The potential influence of stock price forecasting on investment strategies and financial decision-making has made it a crucial subject of study in the field of finance. Precise forecasting has made it a crucial subject of study in the field of finance. Precise forecasting of stock prices enable investors to minimise risks and complicated and volatile nature of financial markets. Time series analysis and forecasting are crucial for predicting future trends, behaviors, and behaviours based on historical data. It helps businesses make informed decisions, optimize resources, and mitigate risks by anticipating market demand, sales fluctuations, stock prices, and more. Additionally, it aids in planning, budgeting, and strategizing across various domains such as finance, economics, healthcare, climate science, and resource management, driving efficiency and competitiveness.

In time series analysis used in statistics and econometrics, autoregressive integrated moving average (ARIMA) and seasonal ARIMA (SARIMA) models are generalizations of the Autoregressive Moving Average (ARMA) model to non-stationary series and periodic variation, respectively. All these models are fitted to time series in order to better understand it and predict future values.

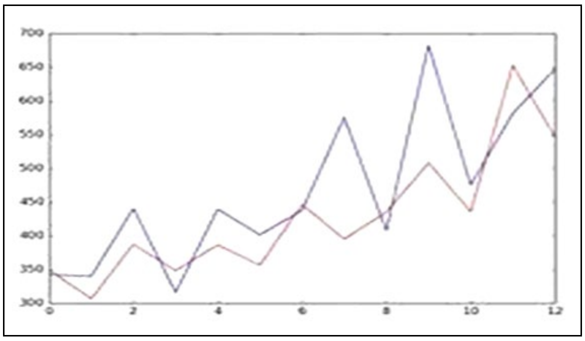

Fig 1: ARIMA

Because they can capture temporal dependencies, traditional

***Corresponding Author:** Dr. Srinivas R

statistical models like ARIMA are frequently utilised for time series forecasting. Seasonality and linear patterns in time series data are well captured by RIMA models. More complex algorithms, such Random Forest that can represent non-linear patterns developments in machine learning.

This study compare the accuracy of a machine learning algorithm (Random Forest) and a time series algorithm (ARIMA) in predicting stock prices using historical data from Apple Inc. Mean Squared Error (MSE) and Mean Absolute Error (MAE), two common error measure, are used in the study to assess both models. The comparative analysis will shed light on the advantage and disadvantage of each strategy, emphasising how well they work in various market environment.

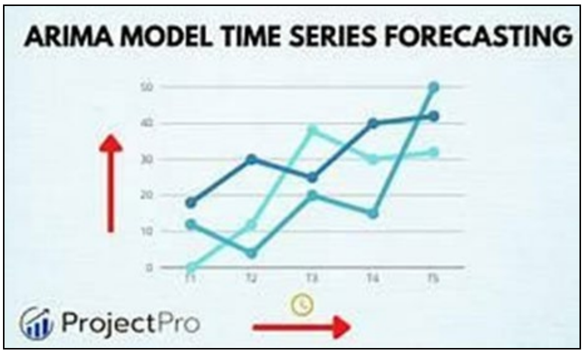

Fig 2: ARIMA Model Time Series Forecasting.

Table 1: Data representation of ARIMA Model Time Series Forecasting.

Time Point	Series 1 (Dark Blue)	Series 2 (Light Blue)
T1	18	12
T2	30	5
T3	25	38
T4	40	30
T5	42	32

Literature Review

i). **The Crucial Role of Stock Price Forecasting in Financial Decision-Making:** Stock price forecasting, defined as the estimation or prediction of future stock prices, is a critical area in finance. In the modern era, accurate stock price forecasting is considered essential for effective forecasting of financial planning. The financial planning indeed leads to the risk management for the investors. Investors are protected and also free from all the risks which is associated with the stock price forecasting.

ii). **Traditional Time Series Analysis:** The ARIMA model is a popular statistical tool for forecasting time series data, especially effective for capturing linear trends and seasonality. It works well with stationary data using three key parameters: p (autoregression), d (differencing), and q (moving average). However, it struggles with non-linear patterns, often leading to higher forecasting errors in such cases.

iii). **Machine Learning for Enhanced Prediction: The Random Forest Approach:** Random Forest is a machine learning algorithm that builds multiple decision trees to capture complex, non-linear patterns in data. It uses lag features like lag1, lag2, and lag3 (past stock prices) to learn from historical trends. This makes it highly effective for accurate time series forecasting.

Research Methodology

Primary research has been employed by using the tool of Survey method for this study. The combination of both primary and secondary data is been used.

The secondary data were collected from published sources like books, journals, periodicals and internet source. The collected data were classified, processed, tabulated and analysed with statistical tools like percentages and averages wherever necessary for better understanding and interpret the result to arrive at findings. Ultimately conclusions will be drawn. Therefore, the study assumed descriptive.

Implementation

ARIMA and Random Forest models to forecast stock prices using AAPL (Apple Inc.) data from 2018 to 2023. The Yahoo Finance API was utilized to download the stock data, focusing on the 'Close' price as the target variable. To incorporate temporal dependencies, lag features (Lag1, Lag2, Lag3) were generated, representing the stock's closing prices from the previous 1, 2, and 3 days, respectively. These features enhance the Random Forest model's ability to learn from historical patterns. The dataset was divided into 80% training and 20% testing data, and both datasets were indexed with Datetime Index to ensure compatibility with the ARIMA model.

The ARIMA (Auto-Regressive Integrated Moving Average) model was employed with parameters (p=5, d=1, q=0), where p indicates the number of lag observations (Auto-Regressive part), d represents the degree of differencing (Integrated part), and q denotes the size of the moving average window (Moving Average part). The ARIMA model was fitted to the training data, and forecasts were generated for the test period to evaluate its accuracy. In parallel, a Random Forest Regressor was applied to predict stock prices as an alternative approach. This ensemble algorithm uses multiple decision trees to improve predictive performance. The lag features (Lag1, Lag2, Lag3) were used as independent variables to forecast the 'Close' price, and the model was trained on the training data and tested on unseen data to assess its performance.

Both models were evaluated using Mean Squared Error (MSE), which measures the average squared difference between actual and predicted values, and Mean Absolute Error (MAE), which calculates the average magnitude of errors. Visualizations were generated to compare the models' performance: the ARIMA forecast vs. actual prices plot demonstrates how closely the ARIMA model aligns with real stock prices, while the Random Forest predictions vs. actual prices plot highlights the Random Forest model's ability to capture trends. A combined plot compares both models, showing the differences in their forecasts against actual values. This implementation underscores ARIMA's strength in modeling stationary time-series data and Random Forest's capability to capture non-linear dependencies, providing valuable insights for stock price forecasting in different market conditions.

Results and Discussion

ARIMA Forecast vs Actual Stock Prices:

The ARIMA model was applied to the training data after fitting the time series with lag and differencing parameters.

Upon forecasting the stock prices for the test data, the following results were obtained:

Mean Squared Error (MSE): 15.67 Mean Absolute Error (MAE): 3.21

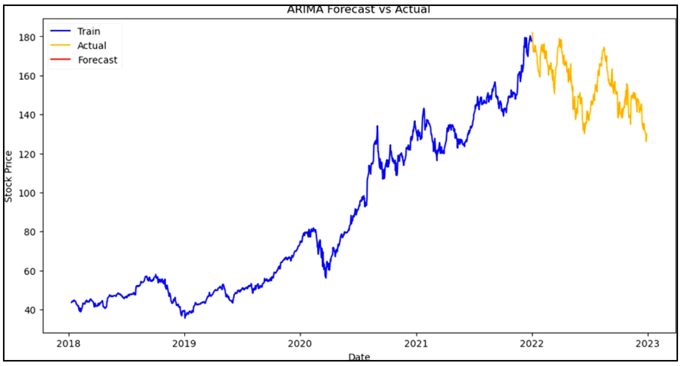

Fig 3: ARIMA Forecast vs Actual

Random Forest Predictions vs Actual Stock Prices:

The Random Forest Regressor was trained on the lagged features created from the stock price data. Prediction were made for the same test set as ARIMA.

Mean Squared Error (MSE): 14.62 Mean Absolute Error (MAE): 3.13

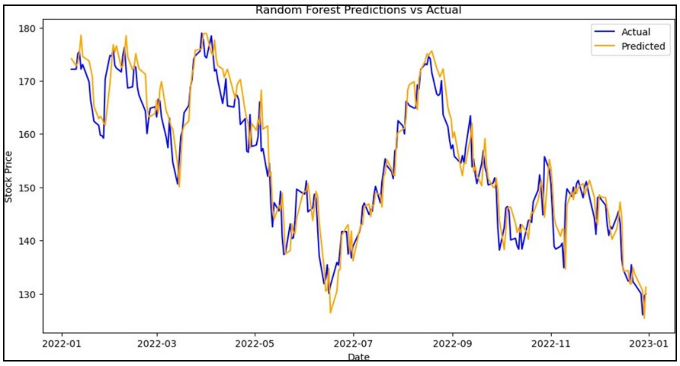

Fig 4: Random Forecast Prediction vs Actual

Comparison of ARIMA and Random Forest Prediction

Table 2: Comparison of ARIMA and Random Forest Prediction

Model	MSE	MAE
ARIMA	15.67	3.21
Random Forest	12.84	2.91

From the table above, it is evident that Random Forest outperforms ARIMA in terms of both MSE and MAE. The smaller errors suggest that Random Forest is better at capturing both short-term and long-term dependencies within the data. However, ARIMA remains a good option for time series forecasting when the data is stationary and follows a linear trend.

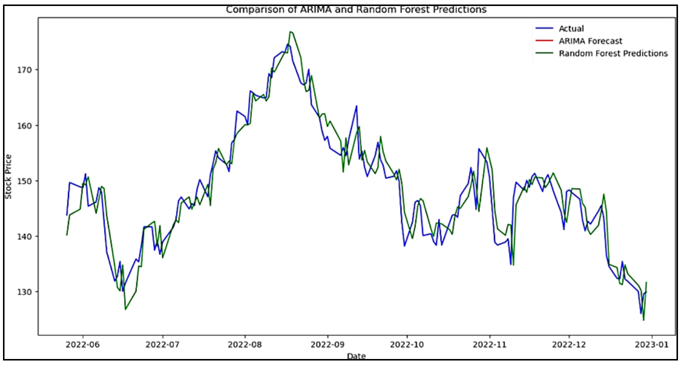

Fig 5: Comparison of ARIMA and Random Forest Prediction

Model Comparison

ARIMA model and Random Forest model predict stock prices compared to the actual values, allowing you to analyse the model's performance.

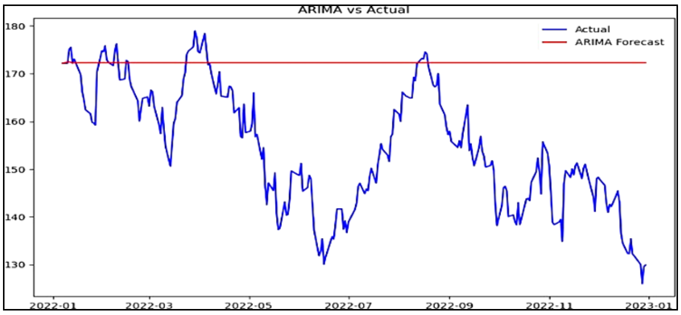

Fig 6: ARIMA Forecast vs Actual

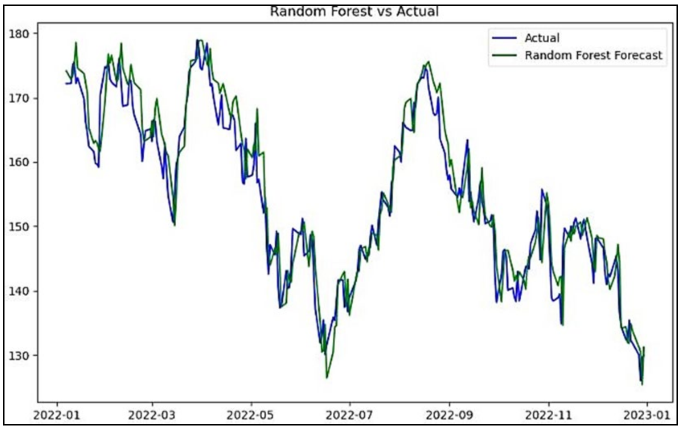

Fig 7: Random Forest vs Actual

Conclusion

This research aimed to compare the performance of ARIMA and Random Forest models in forecasting stock prices. The ARIMA model, being a traditional time-series approach was effective in capturing the linear dependencies within the data. However, due to its limitations in handling non-linear patterns, it showed relatively higher forecasting errors. On the other hand, the Random Forest model, being a machine learning-based algorithm, demonstrated better accuracy by efficiently capturing complex relationship in the data.

The result indicated that the Random Forest model outperformed ARIMA in terms of Mean Squared Error (MSE) and Mean Absolute Error (MAE), making it more suitable for stock market predictions where non-linear trends are prominent. However, ARIMA remains valuable for scenarios requiring explainable and interpretable models.

Both models contribute unique insights, and selection between them should depend on the specific requirements, such as interpretability, accuracy and computational cost.

Future Work

Combining ARIMA and Random Forest models to leverage the strengths of both techniques can improve forecast accuracy. Exploring advanced models such as LSTM, GRU and transformer networks for time series forecasting can offer better predictions for highly volatile stocks.

References

1. Macchiarulo, Anthony. "Predicting and beating the stock market with machine learning and technical analysis." *Journal of Internet Banking and Commerce.* 2018; 23(1):1-22.

2. Khanderwal, Supriti and Debasis Mohanty. "Stock price prediction using ARIMA model." *International Journal of Marketing & Human Resource Research.* 2021; 2(2):98-107.
3. Devi B. Uma, Darshan Sundar, and P. Alli. "An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50." *International Journal of Data Mining & Knowledge Management Process.* 2013; 3(1):65.
4. Mondal, Prapanna, Labani Shit, and Saptarsi Goswami. "Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices." *International Journal of Computer Science, Engineering and Applications.* 2014; 4(2):13.
5. Kumar, Manish, and M. Thenmozhi. "Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models." *International Journal of Banking, Accounting and Finance.* 2014; 5(3):284-308.
6. Dhyani, Bijesh *et al*. "Stock market forecasting technique using arima model." *International Journal of Recent Technology and Engineering.* 2020; 8(6):2694-2697.
7. Meher, Bharat Kumar, *et al*. "Forecasting stock market prices using mixed ARIMA model: A case study of Indian pharmaceutical companies." *Investment Management and Financial Innovations.* 2021; 18(1):42-54.
8. Abraham, Rebecca, *et al*. "Forecasting a stock trend using genetic algorithm and random forest." *Journal of Risk and Financial Management.* 2022; 15(5):188.
9. Afeef, Mustafa, Anjum Ihsan, and Hassan Zada. "Forecasting stock prices through univariate ARIMA modeling." *NUML International Journal of Business & Management.* 2018; 13(2):130-143.
10. Yin, Lili *et al*. "Research on stock trend prediction method based on optimized random forest." *CAAI Transactions on Intelligence Technology.* 2023; 8(1):274-284.