ISSN: 2231- 4687
Vol.-2 Issue-01, January 2026

International Journal of Management and Economics (IJME)

Stock Price Prediction Using Statistical Modeling

Prof. Yogita M. Sadani[1], Ms. Nikita J. Ware[2]

[1]Assistant Professor, Department of Statistics
Dr. D. Y. Patil. Arts, Commerce, & Science College,
Akurdi, Pune-44
Email Id: ysadani82@gmail.com

[2]Student, Department of Statistics
Dr. D. Y. Patil. Arts, Commerce, & Science College,
Akurdi, Pune-44
Email Id: warenikita71@gmail.com

ABSTRACT:

Accurate prediction of stock market behavior remains a critical challenge due to the inherent volatility and complexity of financial markets. Reliable forecasting models are essential for investors, analysts, and policymakers to support informed decision-making and effective risk management. This study investigates the application of statistical modeling techniques, with a primary focus on the Autoregressive Integrated Moving Average (ARIMA) model, for forecasting stock price movements. Historical stock price data were collected from publicly available financial sources and subjected to systematic preprocessing, including data cleaning, normalization, and transformation, to ensure analytical accuracy.

The study employs both regression analysis and ARIMA modeling to examine patterns in stock prices and to evaluate their predictive capabilities. The regression model provides insights into the linear relationships between price variables, while the ARIMA model captures temporal dependencies within the time series data. Model performance was assessed using standard evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2).

The results indicate that the ARIMA model demonstrates strong predictive accuracy for short-term forecasting, with high explanatory power and relatively low prediction error. These findings highlight the effectiveness of statistical time-series approaches in modeling financial data, particularly when the objective is short-term trend estimation. Although market behavior is influenced by numerous external and unpredictable factors, the study confirms that statistical forecasting models can serve as valuable decision-support tools. Future research may focus on

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44

integrating machine learning and deep learning techniques with traditional statistical models to further enhance forecasting performance under dynamic market conditions.

Keywords: ARIMA model, Stock price forecasting, financial prediction, NIFTY 50, Data preprocessing, Regression analysis, Model evaluation

INTRODUCTION:

Forecasting has long been an area of keen interest in research, with continuous efforts devoted to refining existing models and approaches. Accurate prediction empowers both institutions and individual investors to make informed financial decisions, manage risks, and design effective investment strategies for the future. Among the many forecasting challenges, predicting stock prices is one of the most complex tasks due to the volatile and dynamic behavior of financial markets. For investors, reliable forecasting tools are highly desirable, as they can simplify decision-making, improve returns, and reduce risks, thereby motivating researchers to develop more sophisticated predictive models.

Over the years, numerous techniques have been introduced for stock market prediction. Artificial Neural Networks (ANNs), for example, have gained popularity for their ability to detect intricate data patterns and generalize from unseen information. Recently, hybrid models have been explored to combine the strengths of different approaches. From an artificial intelligence perspective, ANNs dominate, while from a statistical standpoint, ARIMA models are widely recognized. Literature suggests that prediction models can be broadly categorized into two groups: statistical methods and AI-based approaches. Among these, ARIMA has proven particularly effective in financial time series forecasting, especially for short-term predictions, often outperforming even advanced neural network techniques. Its extensive use in economics and finance underscores its reliability.

This paper presents the structured development of ARIMA models for short-term stock price forecasting. Empirical results using real-world financial data highlight the model's strength in providing accurate short-term predictions that can support investors in their decision-making processes.

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44

LITERATURE REVIEW:

Traditional Time Series Models

Adebiyi et al. (2014) demonstrated the effectiveness of the ARIMA model in capturing linear trends within financial data, making it a widely accepted choice for time-series forecasting. Jarrett and Kyper (2011) further advanced this approach by incorporating intervention analysis to study Chinese stock prices, showing how external shocks and market events can significantly influence prediction outcomes. These works confirm that ARIMA remains a valuable tool for situations where financial patterns are stable and relatively consistent over time.

Application of ARIMA in Stock Price Prediction

Several studies have applied ARIMA for forecasting stock market behavior. Empirical analyses using stock data from the New York Stock Exchange (NYSE) and Nigerian Stock Exchange (NSE) revealed that ARIMA can provide reliable short-term predictions that compare favorably with alternative statistical and machine learning approaches. Its simplicity, interpretability, and strong performance in short forecasting horizons make it a competitive model in financial research.

ARIMA with Intervention for Market Disruptions

Research on Chinese stock markets has highlighted the benefits of combining ARIMA with intervention analysis. This hybrid approach enables researchers to model the effects of economic disturbances, policy changes, and other shocks on stock indices. Such analyses provide detailed insights into the dynamic relationship between market fluctuations and external factors, proving useful for both researchers and practitioners who seek to understand how disruptions reshape long-term patterns.

NIFTY 50 Stock Price Forecasting

Projects focusing on India’s NIFTY 50 index have used ARIMA and complementary models for forecasting. By training on four years of historical data, these models allow investors to input specific stocks, quantities, and trading timelines to simulate portfolio returns. The integration of forecasting models with tools such as Google Sheets further enhances usability, helping investors conduct real-time portfolio analysis and make informed trading decisions.

Deep Learning and Natural Language Processing Approaches

Recent advancements in deep learning have broadened the scope of prediction models. Mehtab and Sen (2019) applied Long Short-Term Memory (LSTM) networks alongside sentiment analysis using Twitter data to forecast NIFTY 50 prices. Their work highlights the potential of merging social media sentiment with historical stock data to improve prediction accuracy. By incorporating

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44

Natural Language Processing (NLP), they demonstrated how public opinion could be a meaningful predictor of short-term stock movements.

Convolutional Neural Networks for Financial Forecasting

Another approach involves using Convolutional Neural Networks (CNNs) on multivariate time series. Mehtab and Sen (2020) examined four years of NIFTY 50 stock data and employed CNNs to capture complex non-linear relationships. Their hybrid deep learning framework successfully forecasted closing stock prices for 2019, showing the growing importance of digital technologies in financial analytics.

Portfolio-Oriented Projects

Beginner-friendly studies, such as those conducted with Python libraries like Pandas and Scikit-learn, provide practical insights for data scientists. These projects demonstrate how to download stock data (e.g., Microsoft from Yahoo Finance), clean and preprocess datasets, train machine learning models, and back-test results to forecast daily closing prices. Such portfolio projects serve as valuable resources for learners and practitioners seeking hands-on exposure.

Hybrid Predictive Frameworks

Research by Sen and DattaChaudhuri (2017) proposed a hybrid forecasting system that combines statistical techniques with machine learning methods such as regression, decision trees, support vector machines, and neural networks. Applied to stocks like Tata Steel and Hero MotoCorp, the framework showed how integrating different methods enhances short-term forecasting accuracy.

Sector-Specific Forecasting Applications

Sen and DattaChaudhuri (2017) also extended time series forecasting into the healthcare sector, analyzing data from 2010 to 2016. Their structural decomposition approach separated trend, seasonal, and random components to identify sector-specific behaviors. This illustrates the adaptability of forecasting models beyond stock markets, emphasizing the importance of tailoring predictive methods to industry-specific needs.

ARIMA MODEL

The Autoregressive Integrated Moving Average (ARIMA) model, introduced through the Box–Jenkins methodology, is a widely adopted approach for analyzing and forecasting time-series data. The model combines three components: auto-regression, differencing to achieve stationarity, and moving averages to account for random fluctuations.

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44

In this framework, the current value of a time series is represented as a function of its previous values and past error terms. By adjusting the parameters associated with these components, the ARIMA model effectively captures underlying temporal structures and provides reliable short-term forecasts. Due to its balance between simplicity and predictive capability, ARIMA remains a preferred choice in financial time-series analysis.

Mathematically, the ARIMA model expresses the current value of a time series as a linear function of its past values and past forecast errors:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots\dots\dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots\dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Where:

- y_t is the value at time t.
- c is a constant term.
- $\phi_1, \phi_2, \dots\dots\dots, \phi_p$ are the parameters of the autoregressive part of the model.
- $\theta_1, \theta_2, \dots\dots, \theta_q$ are the parameters of the moving average part of the model.
- ϵ_t is the error term at time t.

The ARIMA framework consists of three main components:

- **Autoregressive (AR):** Incorporates past values of the series to explain the current observation.
- **Integrated (I):** Applies differencing to transform non-stationary series into stationary ones.
- **Moving Average (MA):** Uses past forecast errors to refine predictions of current values.

METHODOLOGY

1. Data Collection

The historical stock data for the NIFTY 50 index was collected from Yahoo Finance, covering the period from January 1, 2020, to September 14, 2024.

2. Data Preprocessing

The collected data was imported into Excel for initial preprocessing steps:

- Sorting: Data was sorted to ensure chronological order.
- Removing Duplicates: Any duplicate entries were removed to maintain data integrity.
- Data Cleaning: Non-relevant columns were removed, and necessary conversions (e.g., date formats) were performed to make the data suitable for analysis.

3. Model Development

Two models were applied to the cleaned dataset using R software:

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44

- Regression Model: Applied to establish a baseline for forecasting stock prices.
- ARIMA Model: Developed to capture and forecast the linear combination of past values and errors.

Result:

1. Regression Model

Price = -20.3598 - 0.6786 (Open) + 0.9272 (High) + 0.7522 (Low) + 0.0084 (Volume)

R square=0.99982

2. ARIMA Model

Actual prices vs forecasted prices

Dates	Actual Price	Forecasted Price
16-09-2024	25,383.75	25,422.66
17-09-2024	25,418.55	25,433.91
18-09-2024	25,377.55	25,445.16
19-09-2024	25,415.80	25,456.41
20-09-2024	25,790.95	25,467.66
23-09-2024	25,939.05	25,501.42
24-09-2024	25,940.40	25,512.67
25-09-2024	26,004.15	25,523.92
26-09-2024	26,216.05	25,535.18
27-09-2024	26,178.95	25,546.43
30-09-2024	25,810.85	25,580.18
01-10-2024	25,796.90	25,591.44
03-10-2024	25,250.10	25,613.94
04-10-2024	25,014.60	25,625.19
07-10-2024	24,795.75	25,658.95
08-10-2024	25,013.15	25,670.20
09-10-2024	24,981.95	25,681.45
10-10-2024	24,998.45	25,692.70
11-10-2024	24,964.25	25,703.95
14-10-2024	25,127.95	25,737.71
15-10-2024	25,057.35	25,748.96
16-10-2024	24,971.30	25,760.21
17-10-2024	24,749.85	25,771.47
18-10-2024	24,854.05	25,782.72
21-10-2024	24,781.10	25,816.47
22-10-2024	24,472.10	25,827.73

Dates	Actual Price	Forecasted Price
23-10-2024	24,435.50	25,838.98
24-10-2024	24,399.40	25,850.23
25-10-2024	24,180.80	25,861.48
28-10-2024	24,339.15	25,895.24
29-10-2024	24,466.85	25,906.49
30-10-2024	24,340.85	25,917.74
31-10-2024	24,205.35	25,928.99
01-11-2024	24,304.35	25,940.25
04-11-2024	23,995.35	25,974.00
05-11-2024	24,213.30	25,985.25
06-11-2024	24,484.05	25,996.51
07-11-2024	24,199.35	26,007.76
08-11-2024	24,148.20	26,019.01
11-11-2024	24,141.30	26,052.77
12-11-2024	23,883.45	26,064.02
13-11-2024	23,559.05	26,075.27
14-11-2024	23,532.70	26,086.52
18-11-2024	23,453.80	26,131.53
19-11-2024	23,518.50	26,142.78
21-11-2024	23,349.90	26,165.28
22-11-2024	23,907.25	26,176.54
25-11-2024	24,221.90	26,210.29
26-11-2024	24,194.50	26,221.54
27-11-2024	24,274.90	26,232.80
28-11-2024	23,914.15	26,244.05
29-11-2024	24,131.10	26,255.30

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44

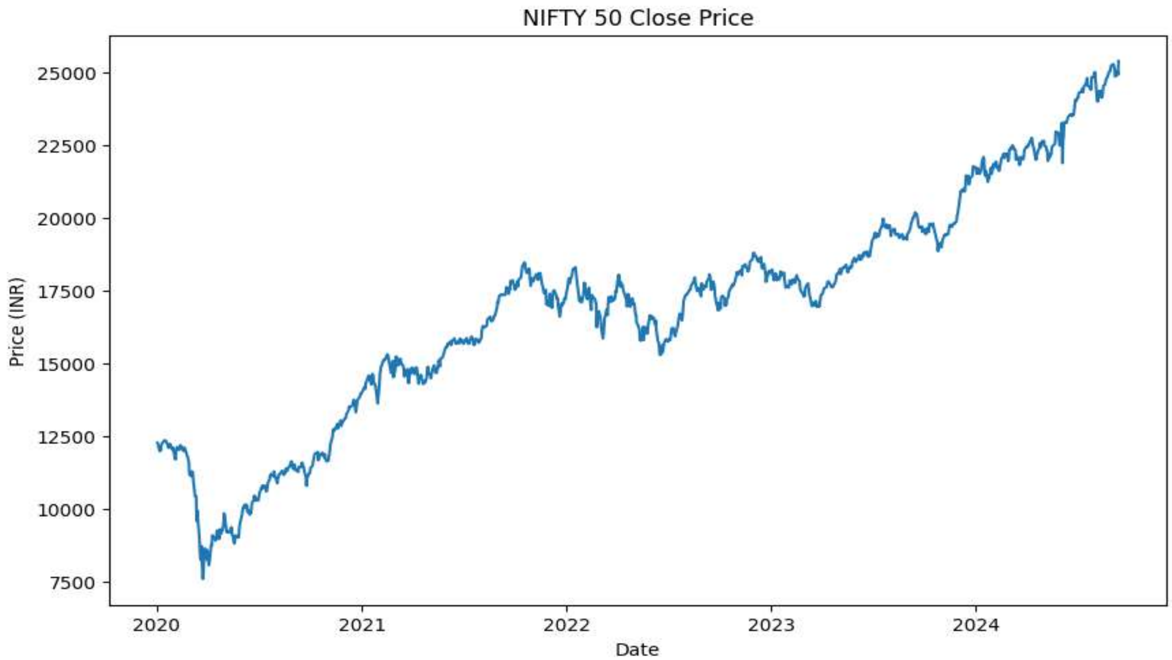

Fig 1: Time series plot of the NIFTY 50 Close Price over time

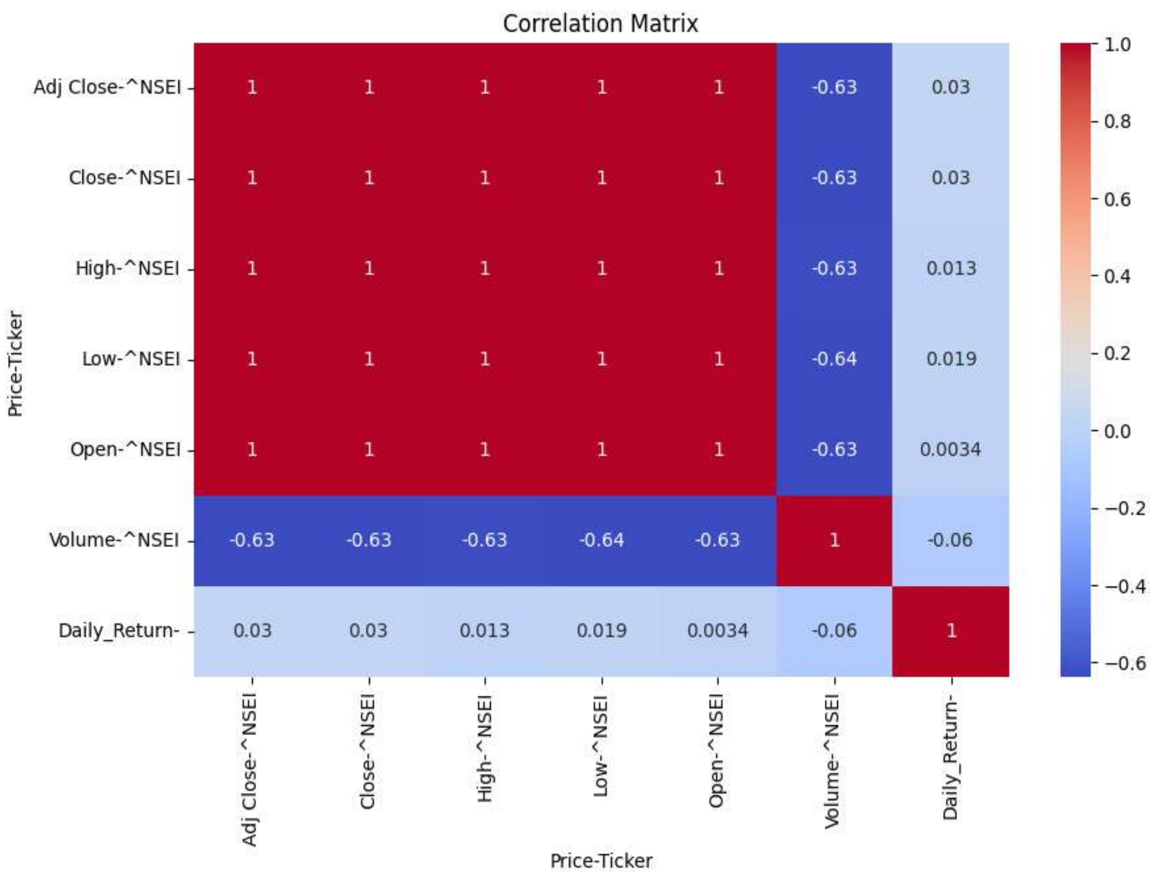

Fig 2: Correlation matrix

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44

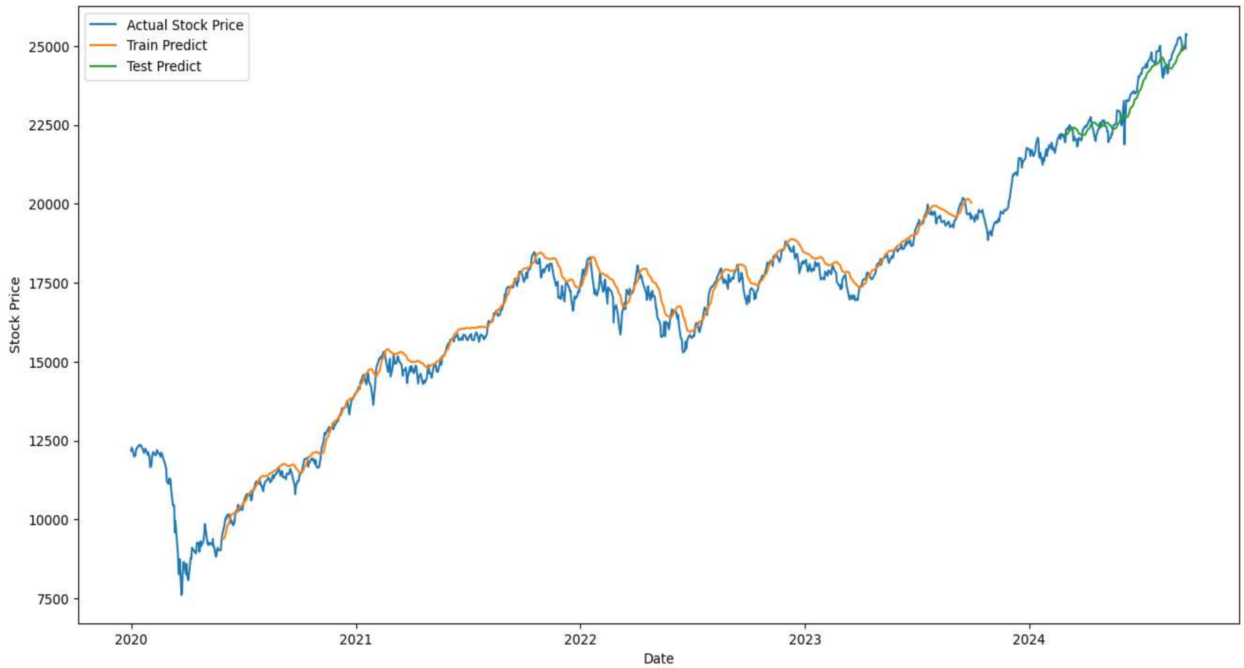

Fig 3: Actual stock price along with predicted values for training and test data.

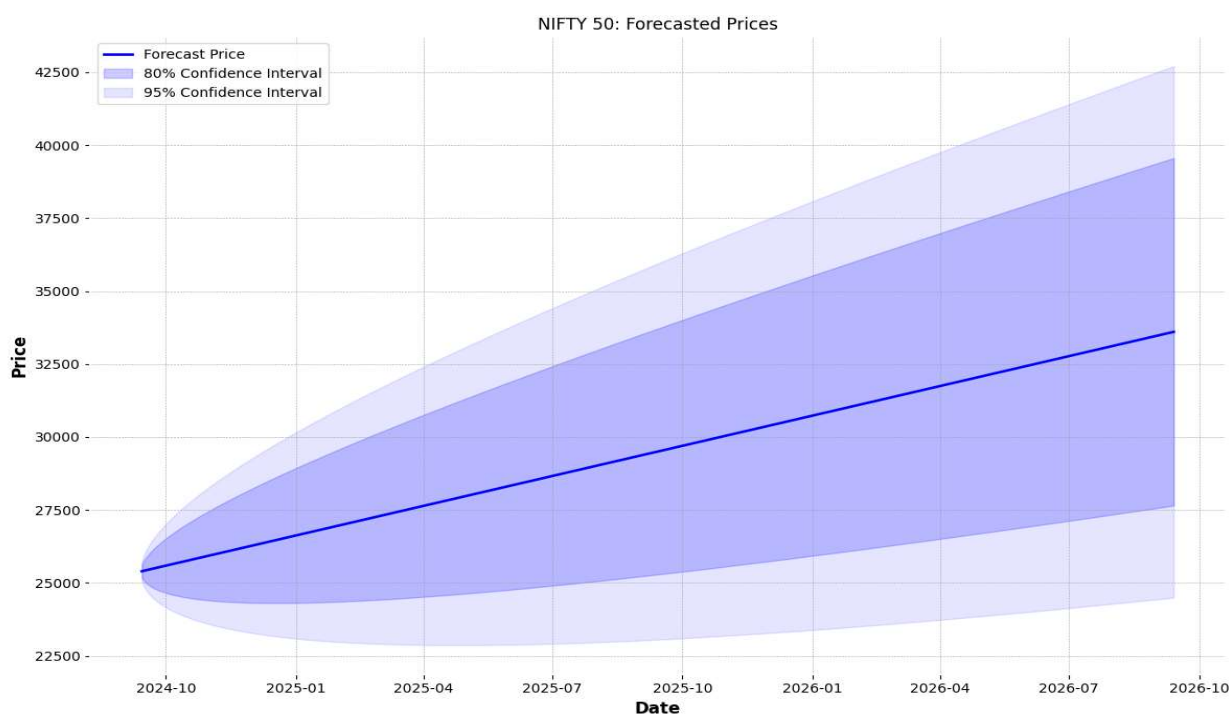

Fig 4: forecasted prices for NIFTY 50along with confidence intervals for next two year

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44

RESULTS:

To evaluate the performance of our ARIMA model for stock price forecasting, we employed several key metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2).

- **Mean Absolute Error (MAE)**: 136.7152
- **Mean Squared Error (MSE)**: 40041.7719
- **Root Mean Squared Error (RMSE)**: 200.1044
- **R-squared (R^2)**: 0.9973

CONCLUSION:

The findings of this study demonstrate that statistical forecasting methods, particularly the ARIMA model, are capable of delivering accurate short-term predictions for stock price movements. The evaluation metrics indicate strong model performance, confirming the suitability of ARIMA for financial time-series analysis. While the model shows high predictive accuracy, it is important to recognize that financial markets are influenced by numerous external and unpredictable factors. Therefore, forecasting models should be used as decision-support tools rather than definitive predictors. Future research may explore hybrid frameworks that integrate statistical techniques with machine learning and deep learning approaches to further enhance forecasting accuracy under dynamic market conditions.

REFERENCES:

1. Adebiyi, A., Adewumi, O., and Ayo, C. K. Ayo: Stock Price Prediction Using the ARIMA Model. In: Proc. of the Int. Conf. on Computer Modelling and Simulation, Cambridge, UK, pp. 105 - 111 (2014)
2. https://www.researchgate.net/publication/261179224_Stock_price_prediction_using_the_ARIMA_model
3. Jarrett, J. E. and Kyper, E.: ARIMA Modelling with Intervention to Forecast and Analyze Chinese Stock Prices. Int. Journal of Engineering Business Management, 3(3), 53 – 58 (2011)
4. Kumar-Dharm/NIFTY50_Stock_Price_Prediction: This project aims to explore the significance of Nifty50, detailing data sources. Showcase model accuracy, a user-friendly

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44

ISSN: 2231- 4687
Vol.-2 Issue-01, January 2026

webpage interface, and profit/loss predictions based on stock details. Highlight real-world applications and outline future project plans.

5. Mehtab, S. and Sen, J.: A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing. In: Proceedings of the 7th Int. Conf. on Business Analytics and Intelligence, Bangalore, India, December 5 – 7 (2019)
6. Mehtab, S. and Sen, J.: Stock Price Prediction Using Convolutional Neural Network on a Multivariate Time Series. In: Proc. of the 3rd Nat. Conf. on Machine Learning and Artificial Intelligence (NCMLAI), New Delhi, INDIA (2020)
7. Portfolio Project: Predicting Stock Prices Using Pandas and Scikit-learn – Dataquest
8. Sen, J. and DattaChaudhuri, T.: A Robust Predictive Model for Stock Price Forecasting. In: Proc. of the 5th Int. Conf. on Business Analytics and Intelligence, Bangalore, India, December 11-13 (2017)
9. Sen, J. and DattaChaudhuri, T.: A Time Series Analysis-Based Forecasting Framework for the Indian Healthcare Sector. Journal of Insurance and Financial Management, 3(1), 66-94 (2017)
10. Yahoo Finance Website: http://in.finance.yahoo.com

1. Prof.Yogita M. Sadani, Assistant Professor, Dept. of Statistics, Dr. D.Y. Patil ACS College, Akurdi, Pune – 44
2. Ms. Nikita Ware, Student, Dr. D.Y. Patil ACS College, Akurdi, Pune - 44