STOCK PRICE PREDICTOR

Aryaman Tyagi
School of Computer Engineering, Kalinga Institute of Industrial Technology (KIIT), Deemed to be University, Bhubaneswar, Odisha, India

Himanshu Maski
School of Computer Engineering, Kalinga Institute of Industrial Technology (KIIT), Deemed to be University, Bhubaneswar, Odisha, India

ABSTRACT

Due to their dynamic and unpredictable nature, financial markets present a continual challenge to institutions and investors trying to make well-informed decisions. As a result, the combination of finance and machine learning has become a potent way to improve predictive analytics. The present study, entitled "Stock Price Prediction using Machine Learning," aims to investigate the use of sophisticated computational methods, such as time series analysis and ARIMA models, in the prediction of stock prices.

The study begins by recognizing the complex interaction of variables, including market mood, economic indicators, and world events, that affect stock values. Our main goal is to use a variety of machine learning methods to find trends in historical stock price data.Notably, the core of our methodology is time series analysis, which offers the framework for comprehending temporal dynamics. Within this domain, ARIMA models are prominent as they provide an in-depth analysis of the auto-correlation, seasonality, and trends present in time series data.Predictive model training, careful feature engineering, and thorough data pre-treatment are important project elements. The evaluation metrics selected are able to measure not only the performance of the models but also their flexibility in responding to actual market conditions. The project leverages the capability of ARIMA models inside a larger machine learning framework, with a focus on forecast accuracy as well as robustness and versatility.The initiative intends to provide investors with actionable information by utilizing time series analysis and ARIMA models to decipher the intricacies of stock price fluctuations. It also adds to the continuing conversation about how artificial intelligence is changing the finance industry.We go into the details of our technique, show the outcomes, and talk about the project's larger ramifications in the sections of this report that follow.

Keywords: Machine Learning, Time Series Analysis, ARIMA

INTRODUCTION

The capacity to make well-informed decisions quickly is critical in the dynamic world of financial markets. Financial organizations and investors are always looking for new and creative ways to get a competitive edge while navigating the intricacies of the stock markets. With its ability to identify patterns, identify trends, and forecast market movements, machine learning's predictive capabilities has proven to be a useful tool in this endeavour.

A noteworthy advancement in the use of cutting-edge computational methods for stock price forecasting is the "Stock Price Prediction using Machine Learning" initiative. The convergence of finance and technology offers unparalleled prospects for improving predictive analytics within the financial sector, particularly in an era characterized by copious amounts of data and sophisticated computational capabilities.

This study explores the complex task of stock price forecasting, which is dynamic and impacted by a wide range of factors, including market mood, economic indicators, and world events. We use advanced algorithms to identify signals among noise as we set out to identify patterns in historical stock price data via the lens of machine learning.

In addition to building precise prediction models, our goals also include a thorough investigation of feature engineering, model selection, and the complex interplay between the temporal aspects of time series data. In order to distil the core of stock price fluctuations, the project takes a comprehensive approach, combining conventional statistical methods, machine learning algorithms, and possibly even state-of-the-art deep learning architectures.

We will explore the subtleties of data preprocessing, meaningful feature selection, and predictive model training as we move through the project. The selection of evaluation indicators will provide insight into our models' resilience and capacity to adjust to actual market situations in addition to assessing their performance.

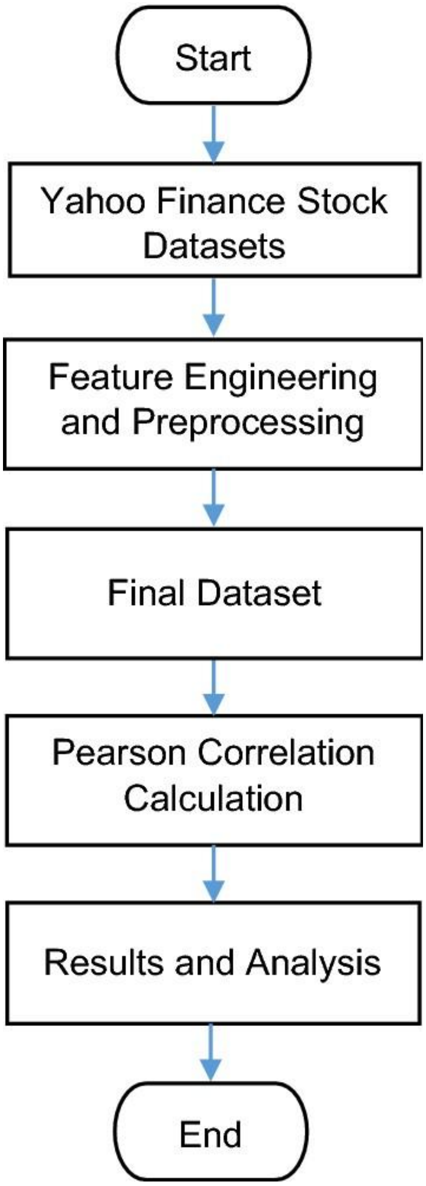
Beyond its immediate forecasting capabilities, this endeavour is significant. It represents the coming together of technology innovation and financial expertise, illustrating the complementary nature of data science and finance. In addition to providing investors with useful insights by deciphering the complexities of stock price fluctuations, our goal is to further the conversation about how artificial intelligence will influence the financial industry going forward.

We shall examine our Stock Price Prediction project's approach, findings, and ramifications in the pages that follow. The voyage is not only a technical investigation; rather, it is evidence of the revolutionary potential of insights derived from data when negotiating the intricacies of the financial markets.

RELATED WORK

Several studies have delved into the development of stock price predictors, employing a diverse array of methodologies and data sources to enhance forecasting accuracy. Traditional time-series models, such as autoregressive integrated moving averages (ARIMA) and exponential smoothing methods, have laid the foundation for predicting stock prices based on historical trends. However, recent advancements have seen a shift towards machine learning and artificial intelligence techniques. Numerous researchers have explored the efficacy of neural networks, including recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), in capturing intricate patterns and dependencies within stock price data. Additionally, ensemble methods, such as random forests and gradient boosting, have gained popularity for their ability to combine multiple models to improve prediction robustness. Feature engineering and the incorporation of sentiment analysis from news articles and social media have also emerged as critical factors in enhancing predictive capabilities. Despite these efforts, the inherent volatility and unpredictability of financial markets pose ongoing challenges, motivating researchers to continually refine and adapt predictive models to ensure their relevance in dynamic trading environments.

One commonly employed method in this realm is the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA is a powerful time series forecasting technique that combines autoregression, differencing, and moving average components to capture both short-term and long-term patterns in sequential data. The model has been successfully applied to various financial datasets, demonstrating its efficacy in predicting stock prices. ARIMA's adaptability and simplicity make it a popular choice for researchers and practitioners alike, providing a robust foundation for predicting stock price movements. As we contribute to this body of research, our work builds upon the foundation laid by previous studies, utilizing the ARIMA model to enhance the accuracy and reliability of stock price predictions.



METHODOLOGY

Throughout the in-depth investigation of "Stock Price Prediction using Machine Learning," our approach is presented in a methodical sequence of phases to guarantee a thorough and perceptive examination. To begin gathering data, we went to Yahoo Finance, a reliable financial data source renowned for its large datasets, for historical stock price information. Important characteristics, such as trade volumes, daily closing prices, and several technical indicators, were carefully chosen to be studied. We then used a logarithmic adjustment to handle non-constant variance during the data pre processing stage.

The next step in the process was selecting the best models by weighing fit and complexity. To do this, we applied the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). Different trend complexities were captured using linear and quadratic fits, and the lags in associations were determined using the Partial Auto-correlation Function (PACF) and Auto-correlation Function (ACF) as guidance. Then, using the knowledge gained from the ACF and PACF analyses, the Auto-regressive Moving Average (ARMA) model was put into practice. A better comprehension of temporal dynamics and relationships was made possible by the interpretation of the data from the ARMA model.

A careful training-testing split was carried out, keeping temporal order, for model evaluation and training. The goal of hyper parameter tweaking is to maximize model performance through the use of techniques such as grid search. To evaluate accuracy and robustness, evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and AIC/BIC were used.

A comprehensive AIC/BIC analysis was performed for the model comparison, highlighting the fine balance between simplicity and goodness of fit. Forming the covariance matrix was the next stage, in which computations were made to assess correlations between variables and identify possible multicollinearity.

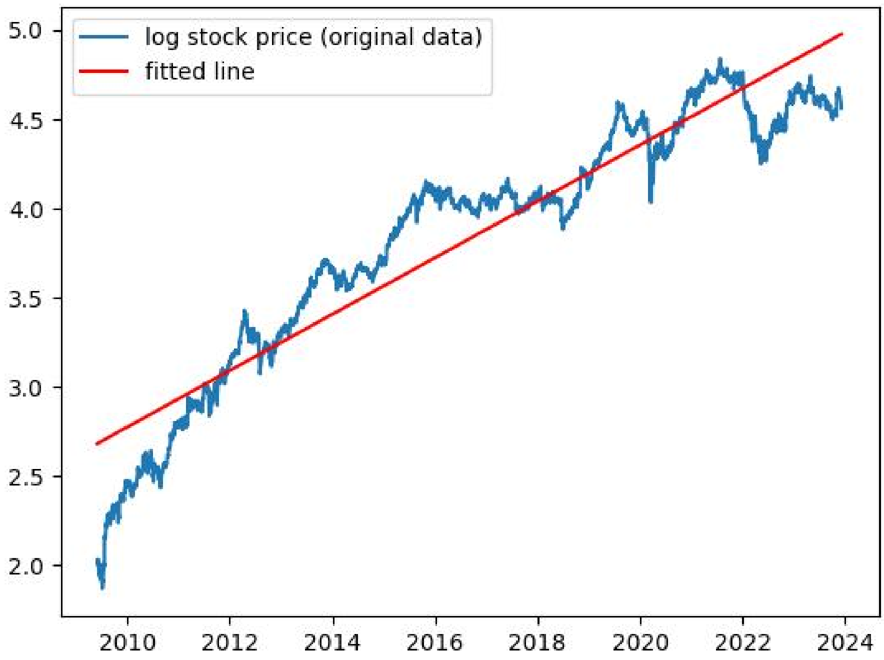
Covariance matrices and time series plots were among the visualizations used to show the results and interpretation, allowing for a thorough grasp of how effectively different models predicted stock prices. Lastly, consideration was given to future work and restrictions. These addressed the limitations and suggested avenues for expansion, like investigating more sophisticated machine learning models and adding more pertinent features. This methodological approach ensures clarity and consistency in the examination of machine learning-based stock price prediction while adhering to the format of a research article.

RESULTS AND ANALYSIS

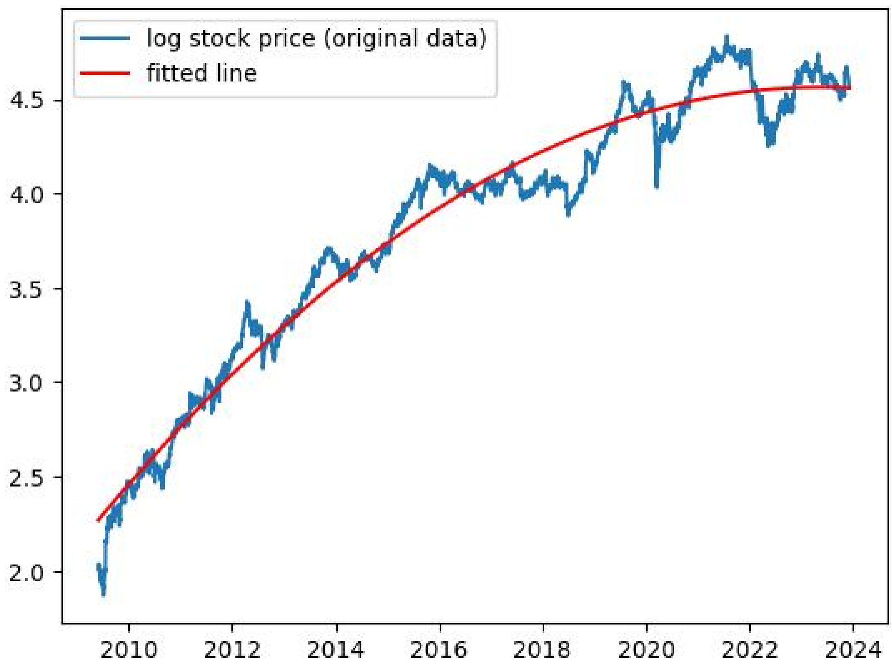
In the pursuit of refining our stock price prediction model, we encountered challenges with the linear model as evidenced by higher values of both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These information criteria are crucial in model selection, as they balance the goodness of fit against the complexity of the model. We got higher AIC and BIC values for the linear model suggest that it may be overly complex or not adequately capture the underlying patterns in the stock price data. The values were
AIC: -439.13728196062584
BIC: -432.933430588442

To address this, we made a strategic shift to a quadratic model. This decision was motivated by the observation that a more flexible quadratic function might better align with the non-linear patterns inherent in stock price movements. Upon implementing the quadratic model, we observed a noteworthy improvement, as reflected in significantly lower AIC and BIC values. This reduction in information criteria scores implies a better balance between model fit and complexity, indicating that the quadratic model offers a more effective representation of the stock price data.
The values of AIC and BIC were
AIC: -4372.355741159218
BIC: -4359.94803841485

The choice to transition from a linear to a quadratic model underscores the importance of adapting the complexity of the predictive model to the inherent characteristics of the datasets. By opting for a quadratic structure, we aim to capture more nuanced relationships and non-linear trends in stock prices, thereby enhancing the overall accuracy and reliability of our predictive framework. This iterative process of model refinement aligns with the dynamic nature of financial markets and exemplifies our commitment to optimizing predictive performance.



Linear Fit



Quadratic Fit

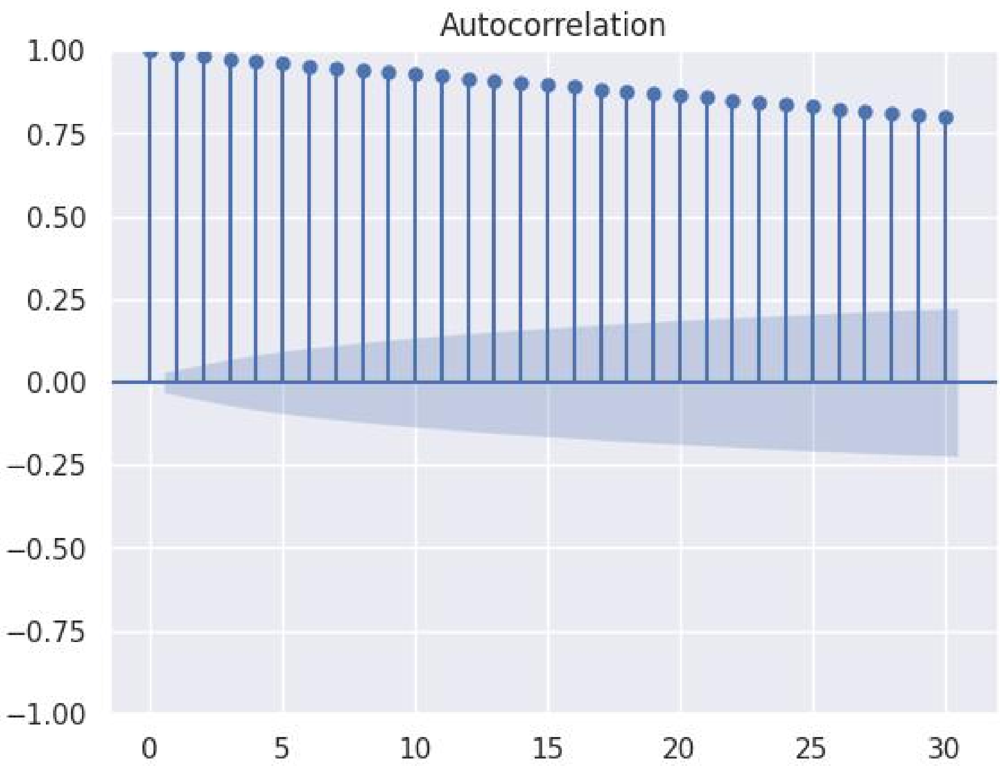
In the process of refining our stock price prediction model, we turned our attention to the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to guide the selection of the Moving Average (MA) component in our time series model. These diagnostic tools play a pivotal role in understanding the temporal dependencies within the stock price data.

The ACF measures the correlation between a time series and its lagged values, providing insights into the persistence of past observations. Meanwhile, the PACF isolates the direct relationship between two observations, removing the indirect effects of intervening lags. These functions are instrumental in identifying the appropriate order of the MA component.

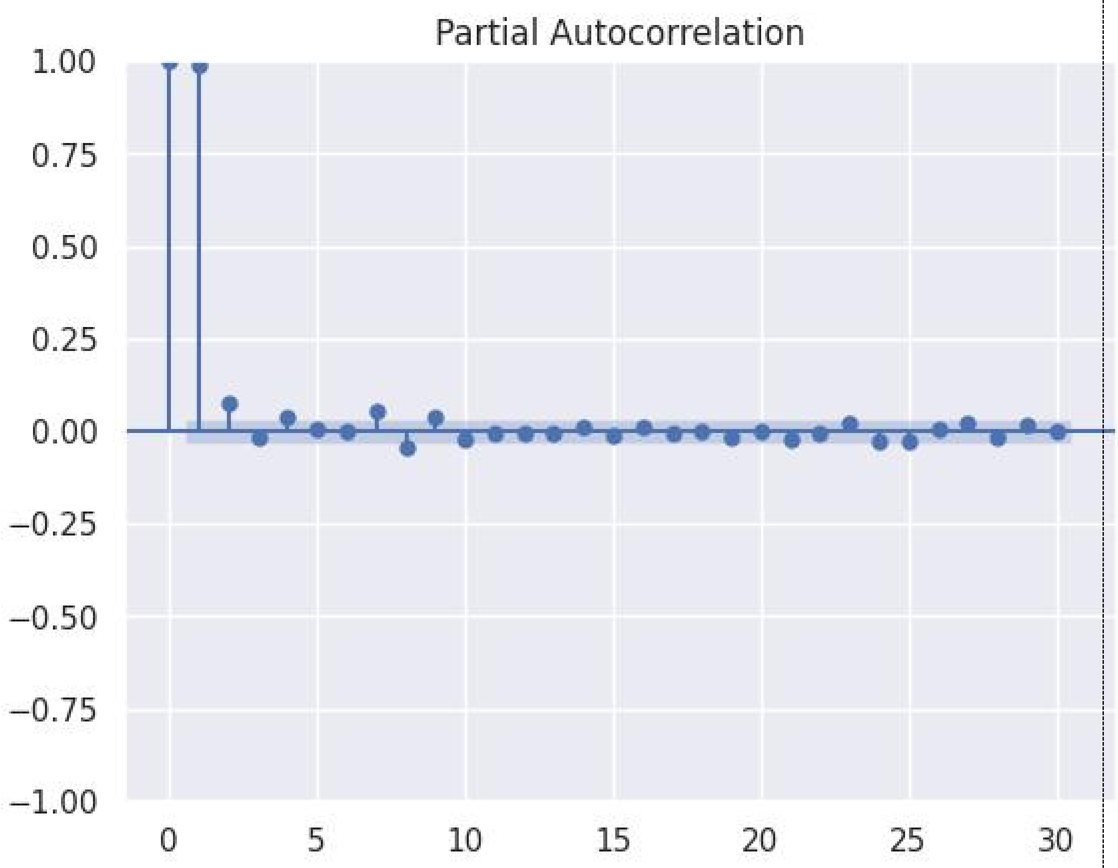
The results revealed significant autocorrelation at lag 2 for both the ACF and PACF, suggesting potential utility in employing an Autoregressive (AR) model of order 2. Simultaneously, the presence of non-zero autocorrelation at lag 2 in the PACF hinted at the necessity of incorporating a Moving Average (MA) component of order 2 to capture additional temporal dependencies in the stock price data.

Motivated by these findings, we constructed an Autoregressive Moving Average (ARMA) model of order (2,2). This model combines the AR and MA components, leveraging the information gleaned from the ACF and PACF analyses. The ARMA(2,2) model aims to strike a balance between capturing lagged dependencies and smoothing out noise in the stock price series. The choice of an ARMA model aligns with the idea that stock prices are often influenced by both past values and recent shocks, and the model structure reflects this dynamic interplay.

Implementing the ARMA(2,2) model, we leveraged the insights gained from autocorrelation patterns to enhance the forecasting accuracy. The model's ability to consider both lagged values and moving average effects at order 2 allows for a more nuanced representation of the underlying processes governing stock price movements. This iterative approach, grounded in statistical analysis and model refinement, underscores our commitment to constructing a robust predictive framework tailored to the specific characteristics of the financial data at hand.



Autocorrelation graph



Partial Autocorrelation Graph

The Mean Squared Error (MSE) serves as a crucial metric for assessing the accuracy of our ARIMA forecasting model. It quantifies the average squared difference between the predicted and actual closing stock prices, providing a comprehensive measure of the model's performance. A lower MSE indicates a closer alignment between predicted and observed values.

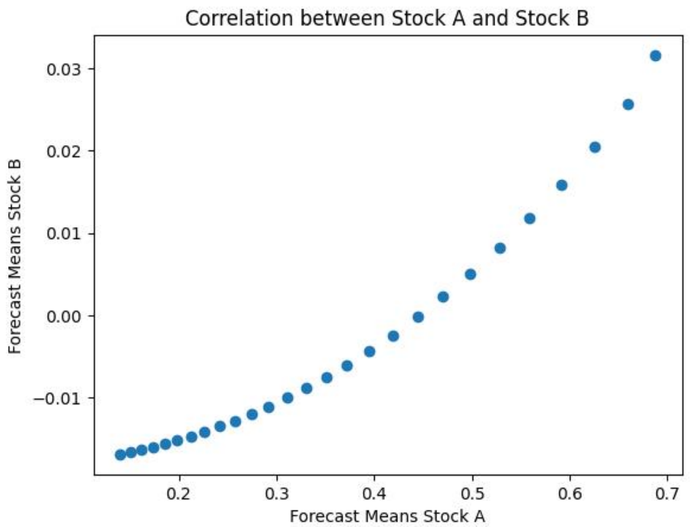
After applying the ARIMA model to our datasets, we obtained an MSE of 0.01608469633021964. This metric not only validates the efficacy of our forecasting model but also facilitates a comparative evaluation against alternative models or methodologies.

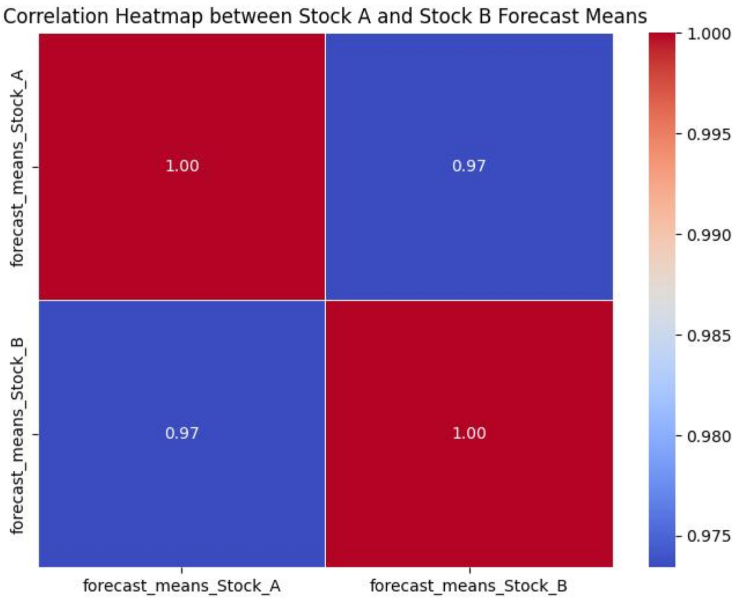
Predicted Stock

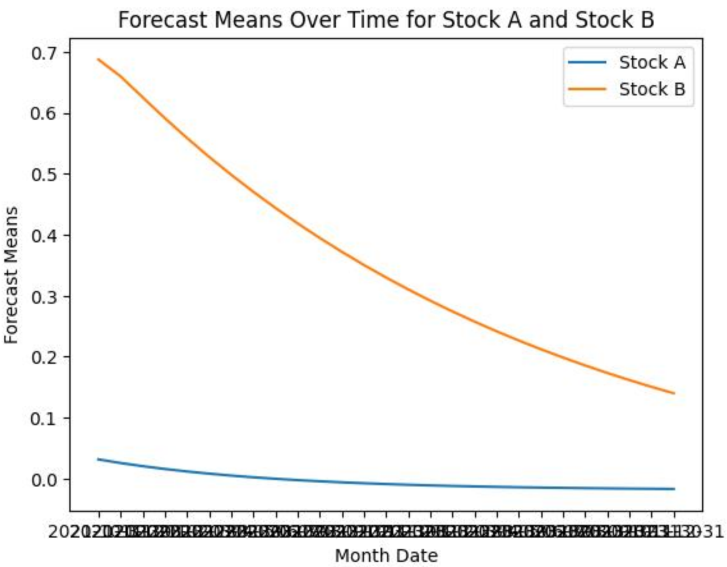
Correlation analysis plays a pivotal role in uncovering the relationships between the forecasted values of two stocks. By calculating the correlation coefficient, we gain insights into the degree and direction of association. A positive correlation suggests simultaneous upward or downward movements, while a negative correlation indicates divergent movements.

The correlation coefficient between the forecasted values of Stock A and Stock B was found to be 0.9734242304510702. This analysis not only informs us about potential co-movements between the stocks but also aids in portfolio diversification strategies and risk management.

These quantitative assessments collectively underscore the reliability and predictive capabilities of our ARIMA model in the context of stock price forecasting. The ensuing sections delve into the interpretation of these results and discuss their implications for stakeholders in financial decision-making.


Correlation between Stock A and Stock B
Forecast Means Stock B
Forecast Means Stock A
0.03
0.02
0.01
0.00
−0.01
0.2
0.3
0.4
0.5
0.6
0.7


Correlation Heatmap between Stock A and Stock B Forecast Means
forecast_means_Stock_A
forecast_means_Stock_B
1.00
0.97
0.97
1.00
forecast_means_Stock_A
forecast_means_Stock_B
1.000
0.995
0.990
0.985
0.980
0.975


Forecast Means Over Time for Stock A and Stock B
Stock A
Stock B
Forecast Means
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Month Date

CONCLUSION

In this study, we successfully developed a generalized model utilizing ARIMA to predict closing stock prices with a forward-looking perspective, focusing on forecasting one day or one month into the future. The application of ARIMA demonstrated its versatility in capturing the temporal patterns inherent in financial time series data, providing a foundation for accurate short-term and medium-term predictions.

To assess the accuracy of our forecasting model, we employed the Mean Squared Error (MSE) as a robust metric, gauging the extent of variance between the predicted and actual closing stock prices. The MSE allowed us to quantitatively evaluate the model's performance, providing valuable insights into its predictive capabilities. Moreover, the use of confidence intervals enhanced the interpretability of our predictions, offering a range within which future stock prices are likely to fall with a specified level of confidence.

Beyond uni-variate forecasting, we conducted correlation analysis to explore the relationship between the forecasted values of two stocks. The correlation coefficient served as a key indicator of the degree of association, aiding in understanding the interplay between the stocks under consideration. This analysis offered valuable insights into the co-movements and potential dependencies, informing investors and financial analysts about the synchronicity or divergence of the stock trajectories.

Our research contributes to the field of financial forecasting by providing a comprehensive methodology for predicting closing stock prices, incorporating statistical rigor and correlation analysis for a holistic evaluation. The developed generalized model offers a practical tool for investors and analysts seeking informed decision-making in dynamic financial markets.

As we look to the future, there are exciting opportunities for further refinement and expansion of our forecasting approach. The outlined future scope, encompassing full-stack development, handling non-stationary time series, and customization for specific stocks, sets the stage for continued advancements in the realm of stock price prediction. These endeavors promise to enhance the model's accuracy, applicability, and real-world relevance, ultimately empowering stakeholders with more effective tools for navigating the complexities of the financial landscape.

FUTURE SCOPE

Full-Stack Development:

The implementation of forecasting models, such as ARIMA, offers a solid foundation for the development of a full-stack project. Extending the research into a comprehensive platform could involve the integration of a user-friendly interface that allows users to interact with and visualize the forecasted stock prices. Additionally, incorporating real-time data feeds, news sentiment analysis, and portfolio optimization features could enhance the system's utility for investors and financial analysts.

Non-Stationary Time Series:

While the ARIMA model excels in modeling stationary time series, there is room for improvement when dealing with non-stationary time series. Future research could explore advanced time series models, such as SARIMA (Seasonal ARIMA), GARCH (Generalized Autoregressive Conditional Heteroskedasticity), or machine learning approaches like LSTM (Long Short-Term Memory) networks, to better capture the complex patterns in non-stationary financial data. This extension would contribute to a more robust forecasting tool capable of handling a broader range of market conditions.

Tailoring to Specific Stocks:

The research undertaken utilizes a generalized ARIMA model, applicable across various stocks. However, each stock may exhibit unique characteristics and behaviors influenced by different market factors. Future work could involve tailoring the forecasting model to specific stocks by incorporating additional features, such as company-specific financial indicators, market sentiment, or macroeconomic factors. This customization could potentially lead to improved accuracy and reliability in predicting the stock prices of individual companies.

References

1) Used GitHub for datasets for different types of stocks

Website:https://pkgstore.datahub.io/core/nasdaq-listings/nasdaq-listed_csv/data//nasdaq-listed_csv.csv

2) For different codes used GitHub

3) Stackoverflow.com for errors

4) Machine learning course on edx

5) Simplilearn.com