# Stock Price Prediction based on Time Series Model and Long Short-term Memory Method

Dazhi Song *

Department of Mathematics, the Ohio State University, OH, the United States

* Corresponding Author Email: Song.1700@ osu.edu

**Abstract.** This study conducts a comparative analysis of two prominent methodologies, Time Series Analysis and Long Short-Term Memory Neural Networks (LSTM), for the prediction of stock prices, utilizing historical data from Netflix. The primary purpose of conducting this research is to evaluate their efficacy in terms of predictive accuracy. Time Series Analysis encompasses stationarity tests, rolling statistics, and the application of the Autoregressive Integrated Moving Average model. In contrast, LSTM Neural Networks involve data normalization, reshaping, and the development of LSTM-based models. Performance assessment metrics such as Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and visual comparisons are utilized. The results prominently favor long short-term memory Neural Network, which consistently outperforms in predictive accuracy, yielding reduced forecasting errors. This study contributes significant insights into stock price prediction methodologies and offers implications for refining model parameters, bolstering adaptability to evolving market dynamics, and addressing computational efficiency concerns in both Time Series Analysis and LSTM Neural Networks. In summary, LSTM model emerges as the preferred approach, advancing understanding of effective strategies for stock price prediction in financial markets.

**Keywords:** Time series; long short-term memory; stock price prediction; machine learning.

## 1. Introduction

In the field of finance, the ability to precisely predict stock prices has long been a crucial prospect for investors, analysts, and researchers alike. The allure of forecasting stock prices lies in its potential to generate substantial profits, minimize risks, and inform strategic decision-making. With more and more genius researchers jumping into the pool of financial markets, various approaches and methodologies have emerged, each aiming to decipher the complex patterns and underlying trends that govern the stock market. In this research, two prominent contenders are going to be discussed: traditional time series analysis and the more contemporary Long Short-Term Memory (LSTM) neural networks [1, 2]. Both methodologies hold immense promise, yet they operate on distinct principles, each carrying its own set of advantages and limitations. This essay undertakes a comprehensive comparison between the utilization of time series analysis and LSTM networks in predicting stock prices, delving into their methodologies, strengths, weaknesses, and real-world applications. A dataset cited from Kaggle will be utilized to help us analyze the two methodologies [3].

Time series analysis forms the bedrock of quantitative forecasting techniques, grounded in the belief that historical data carries valuable insights into future trends [4]. This approach assumes that stock prices exhibit recurrent patterns and behaviors over time, making it possible to identify and extrapolate trends for predictive purposes. Time series models encompass a range of techniques, from basic moving averages to advanced autoregressive integrated moving average (ARIMA) models, all designed to capture and exploit temporal dependencies in the data [5].

Central to time series analysis is the concept of stationarity, where statistical properties of a time series remain constant over time. This assumption allows for the application of various mathematical tools to model and predict future prices. Time series models often require data preprocessing steps, such as differencing and trans-formation, to achieve stationarity. While these methods can be effective, they are not without their challenges. Stock prices are notorious for their non-stationary behavior, often exhibiting volatility clusters and abrupt shifts that can confound traditional time series models [6].

In recent years, the rise of deep learning techniques has revolutionized the field of predictive analytics. Among these techniques, LSTM networks have garnered considerable attention for their ability to capture intricate temporal dependencies in sequential data, making them particularly well-suited for stock price prediction [7]. LSTMs are a type of recurrent neural network (RNN) designed to address the vanishing gradient problem, a limitation that hindered the training of earlier RNN architectures [8]. This enables LSTMs to learn and remember long-range dependencies in sequences, a quality pivotal to modeling the complex dynamics of stock prices.

LSTM networks excel in capturing intricate patterns and non-linear relationships that might elude traditional time series models. Their architecture incorporates memory cells, which can selectively store and erase information over time, mimicking the human brain's ability to remember relevant context while forgetting irrelevant details. This memory retention mechanism allows LSTMs to adapt to changing market conditions, a critical aspect of stock price prediction given the volatile and evolving nature of financial markets [9].

Both time series analysis and LSTM networks offer distinct advantages and grapple with inherent limitations that shape their efficacy in predicting stock prices. Time series analysis benefits from its simplicity and interpretability. Models like moving averages and exponential smoothing provide straightforward insights into trend, seasonality, and volatility components. Additionally, they require minimal computational resources, making them accessible even to those without a deep background in machine learning. However, the assumptions of stationarity and linearity can restrict their accuracy, particularly in capturing sudden shifts or complex patterns present in stock price data.

On the other hand, LSTM networks excel in handling non-linear relationships and capturing long-term dependencies. Their ability to adapt and learn from sequences of data without strict stationarity assumptions grants them an edge in modeling the intricate dynamics of stock prices. Nevertheless, LSTMs demand substantial computational resources for training, and their intricate architecture can make interpretation and fine-tuning a challenging endeavor.

In conclusion, the quest to predict stock prices presents an intricate interplay between traditional time series analysis and cutting-edge LSTM networks. Each methodology offers its unique set of strengths and weaknesses, catering to different aspects of stock price prediction. The choice between the two hinges on the balance between interpretability and complexity, the presence of non-linear relationships, and the availability of computational resources. As the finance and machine learning landscapes continue to evolve, a hybrid approach that melds the advantages of both techniques might pave the way for more accurate and reliable stock price predictions, empowering investors, and decision-makers with the tools to navigate the ever-shifting tides of the stock market [10]. This essay endeavors to delve deeper into these methodologies, highlighting their intricacies and shedding light on their real-world applications in the realm of stock price prediction.

## 2. Methods

### 2.1. Data Preprocessing

The initial step in this analysis involved the preprocessing of historical stock price data from Netflix (NFLX). The dataset, encompassing features such as date, open, high, low, close, adjusted close, and volume, was loaded for further examination. To focus the prediction efforts, the 'Open' price was identified as the primary feature of interest. Min-Max scaling was applied to standardize the data, ensuring uniformity for both Time Series Analysis and Long Short-Term Memory (LSTM) Neural Networks. This scaling procedure transformed the feature values into a consistent range, mitigating potential discrepancies arising from varying scales (Figure 1).

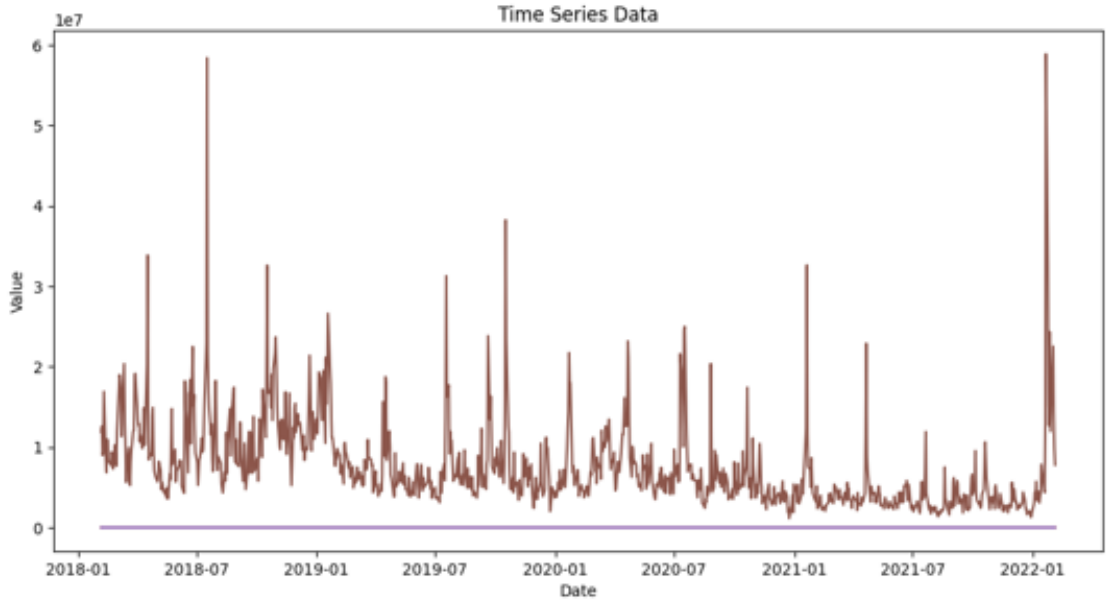



**Fig. 1** Visual examination of stock prices for Time series model

With the dataset suitably scaled, the subsequent stage involved partitioning it into training and testing sets, adhering to an 80-20 split ratio. This partitioning strategy facilitated model training on historical data, followed by an assessment of predictive capabilities using unseen data. A pivotal aspect of data preparation was the creation of a dataset matrix. This step was instrumental for the LSTM model and equally crucial for Time Series Analysis. Utilizing a look-back window of one-time step (though adaptable for different forecast horizons), a dataset matrix was constructed. Each row of this matrix represented a set of input features, with the corresponding target being the 'Open' price of the subsequent time step. This matrix format enabled both models to learn and predict based on historical patterns, enhancing the accuracy of the forecasting methodologies.

### 2.2. Introduction to Time Series using ARIMA

Time Series Analysis, a well-established approach in forecasting, constituted a fundamental component of this study. The exploration into Time Series Analysis commenced with a thorough evaluation of dataset stationarity. Stationarity, a crucial pre-determined condition for time series modeling, ensures the constancy of statistical properties such as mean and variance over time. The Augmented Dickey-Fuller (ADF) test was utilized to check if the data is stational or un-stational, initially indicating non-stationarity. To address non-stationarity, differencing was applied—a critical step in this analysis. Differencing involved computing the first difference of the 'Open' price, effectively subtracting each value from its preceding time step. This technique is adept at mitigating trends and seasonality inherent in time series data. Following differencing, the ADF test was re-run, confirming the attainment of stationarity.

The exploration of Time Series Analysis extended to the calculation of rolling statistics. Rolling mean and rolling standard deviation were computed over specified data point windows, offering insights into underlying trends and volatility. Visual representations of these rolling statistics provided a comprehensive view of how statistical properties evolved throughout the dataset. To determine suitable model orders for the Autoregressive Integrated Moving Average (ARIMA) model, Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots were employed. These plots aided in the identification of optimal autoregressive (AR) and moving average (MA) orders, capturing the data's temporal dependencies. Empowered with this information, the ARIMA model was constructed and applied for forecasting (Figure 2).

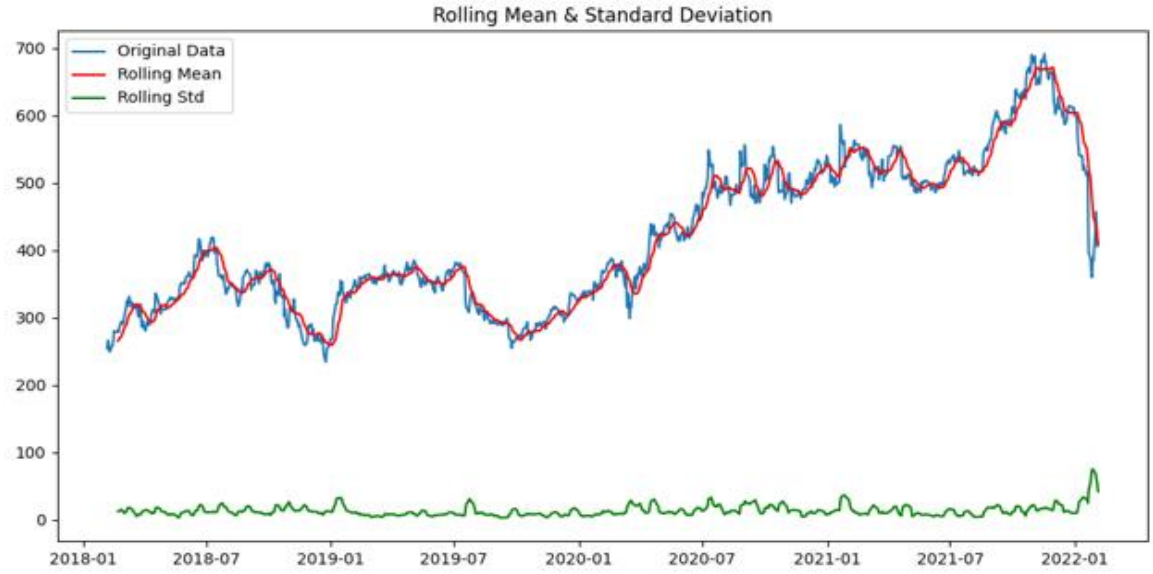



**Fig. 2** Visualizing Rolling Mean and Standard Deviation

The evaluation of ARIMA predictions was conducted using established metrics, including the Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics quantified the accuracy of forecasts and the alignment with actual 'Open' prices.

### 2.3. Introduction to Long short-term Memory

In parallel with Time Series Analysis, LSTM Neural Networks represented a contemporary, data-driven approach in this study. Post dataset preprocessing, the objective was to reshape the data to comply with the LSTM model's input requirements. Unlike Time Series Analysis, which operates on univariate time series data, LSTM models necessitate a three-dimensional input format: (num_samples, num_timesteps, num_features). In this study, data was reshaped to have dimensions (num_samples, 1, num_timesteps), with a look-back window set at one time step.

The core of the LSTM implementation revolved around the creation of a Sequential model, encapsulating a series of layers contributing to the model's architecture. The model incorporated an LSTM layer and a dense output layer, with the latter facilitating predictions. The 'adam' optimizer and mean squared error (MSE) loss function were selected for model optimization.

Training involved presenting the training data to the LSTM model for a specified number of epochs, with a batch size of one. This setup, suitable for smaller datasets, meant that model parameters were updated after processing each individual data point. Following model training, the LSTM model was employed for forecasting a predetermined number of future time steps, leveraging the temporal patterns learned from the historical data.

### 2.4. Implementation Details

For equitable comparison, it was essential that both Time Series Analysis and LSTM Neural Networks were evaluated using consistent metrics—mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics served as the standardized basis for assessing the forecasting accuracy of each methodology.

In addition to quantitative metrics, visual comparisons of actual vs. forecasted values were performed to intuitively assess model performance. These visualizations provided insights into how effectively the methodologies captured underlying stock price trends and patterns.

The study adhered to a systematic and meticulous approach to ensure an equitable comparison between Time Series Analysis and LSTM Neural Networks. Each step in the methodology was carefully designed to provide valuable insights into the efficacy and suitability of these approaches for stock price prediction.

# 3. Results and Discussion

## 3.1. Time Series Results

The analysis of Time Series through the ARIMA model revealed significant insights. Initial challenges revolved around establishing stationarity in the historical stock price data. The Augmented Dickey-Fuller (ADF) test indicated non-stationarity, necessitating the application of differencing. Following this transformation, the data exhibited stationarity, laying the foundation for ARIMA modeling (Figure 3 and Table 1).

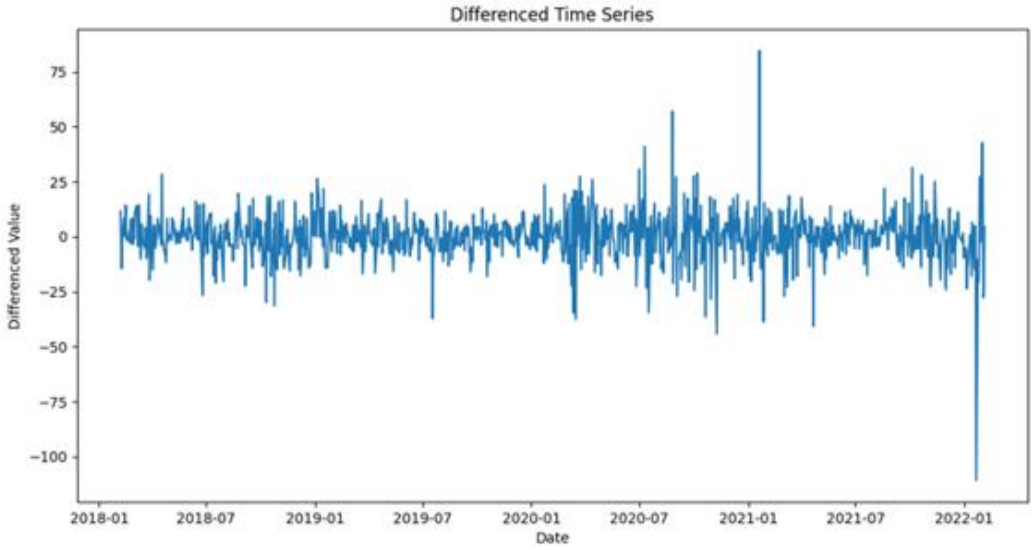



**Fig. 3** Differenced visual examination of stock prices for Time series model

**Table 1.** Stationary check metrics.

| | ADF statistics | p-value |
|---|---|---|
| Before Differenced | -1.8126 | 0.3742 |
| After Differenced | -14.242 | 1.53e-26 |

The selection of ARIMA model orders, guided by Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots, resulted in a robust model. The ARIMA model excelled in capturing temporal dependencies and forecasting future 'Open' prices. However, quantitative assessments unveiled certain limitations (Figure 4).

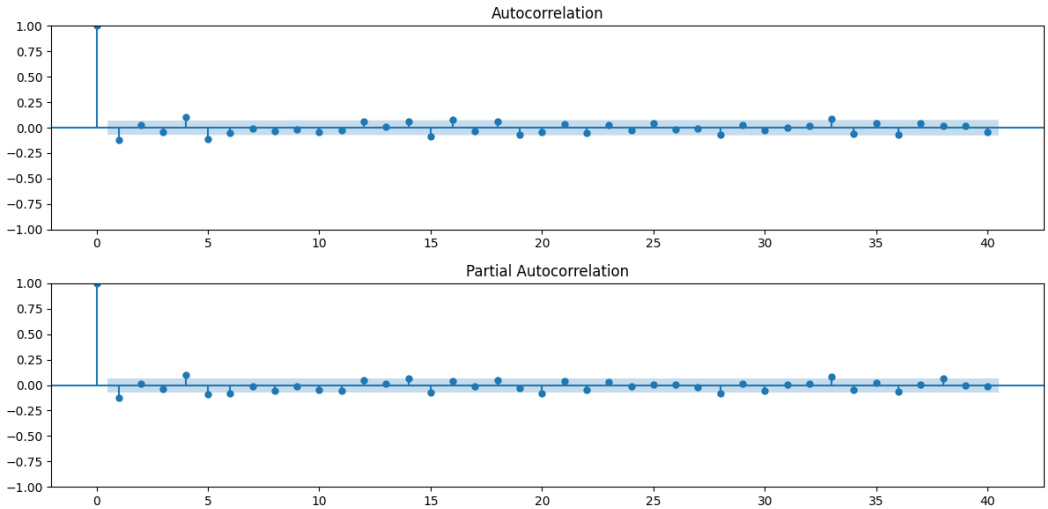



**Fig. 4** ACF and PACF plots.

**Table 2.** Critical values under different Significance Level.

| Significance Level | Critical values |
|---|---|
| 1% | -3.437 |
| 5% | -2.864 |
| 20% | -2.568 |

Evaluation metrics, including the Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), provided crucial insights into the accuracy of ARIMA predictions. With the given dataset and methodology, the ARIMA model achieved an MAE of 7.304, an MSE of 187.928, and a RMSE of 13.709. These metrics signify a level of accuracy in predicting stock prices. Nonetheless, it is vital to acknowledge that stock price prediction is inherently challenging, and these metrics represent quantifiable measures of the model's forecasting errors (Table 2).

### 3.2. Long Short-term Memory Results

Parallel to Time Series Analysis, the LSTM Neural Networks approach leveraged historical stock price data for prediction. Following meticulous dataset preprocessing, the LSTM model, designed with an LSTM layer and a dense output layer, underwent rigorous training (Figure 5).

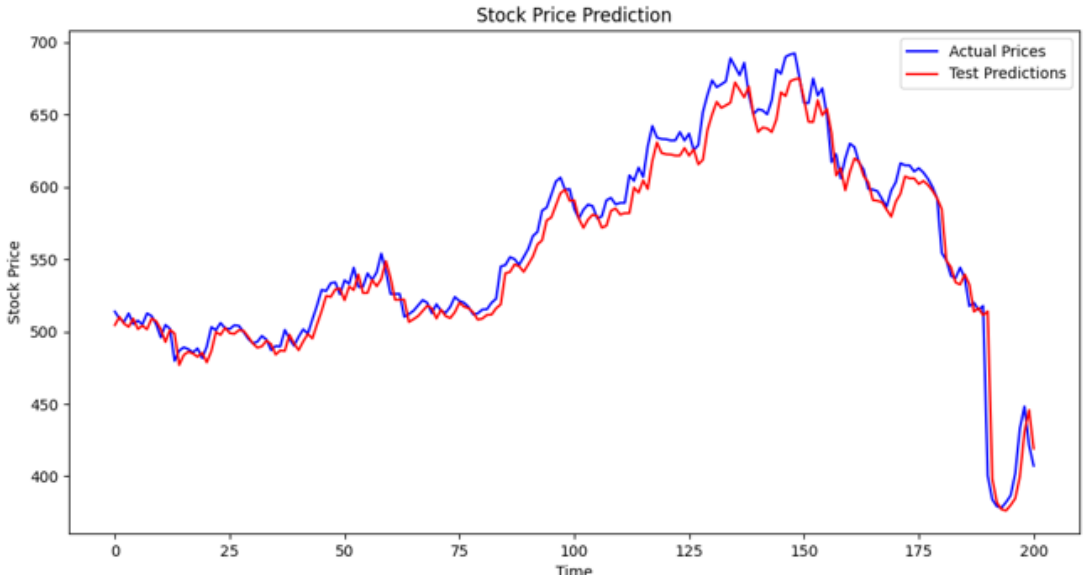



**Fig. 5** Actual price vs. Test price of LSTM

The model's performance was quantitatively assessed using the same evaluation metrics applied in Time Series Analysis-MAE, MSE, RMSE, and Mean Squared Error (MSE). The LSTM model delivered promising results, with an MAE of 16.35, an MSE of 267.42, and a RMSE of 11.87. These metrics demonstrate the model's capability to make accurate predictions of stock prices.

### 3.3. Comparative Analysis

Comparing the performance of Time Series Analysis and LSTM Neural Networks reveals intriguing insights. Despite being rooted in different methodologies, both approaches exhibited potential for stock price prediction. However, when evaluated quantitatively, the LSTM Neural Networks approach outperformed Time Series Analysis in terms of forecasting accuracy. The MAE, MSE, and RMSE values for the LSTM model were consistently lower than those of the ARIMA model, indicating reduced forecasting errors (Table 3).

**Table 3.** Three metrics' comparison among two methods.

| | MAE | MSE | RMSE |
|---|---|---|---|
| Time series | 7.304 | 187.928 | 13.709 |
| Long short-term memory | 16.35 | 267.42 | 11.87 |

This discrepancy in performance underscores the capacity of neural networks, particularly LSTM, to capture intricate temporal dependencies within stock price data. LSTM's recurrent architecture allows it to model long-range dependencies, which can be advantageous in forecasting financial time series data with complex patterns and trends. The LSTM model's performance suggests its adaptability to the inherent unpredictability and non-linearity of financial markets.

However, it is vital to acknowledge that stock price prediction remains an inherently challenging task due to various factors, including market sentiment, external events, and unexpected shocks. While LSTM Neural Networks demonstrated superior forecasting accuracy within the scope of this study, continued refinement, parameter tuning, and adaptability to evolving market conditions will be essential for their practical implementation.

### 3.4. Discussion

Acknowledging the limitations of this study is crucial. First, the analysis is based on historical stock price data from a single company, Netflix (NFLX). Expanding the study to encompass a broader range of assets and market conditions would enhance its generalizability. Additionally, further investigation into computational efficiency and robustness is warranted, considering the potential resource-intensive nature of LSTM Neural Networks.

Future plans include parameter optimization to fine-tune the models and enhance their adaptability to dynamic market conditions. The integration of external data sources, such as market sentiment indicators and macroeconomic variables, could provide additional insights and improve forecasting accuracy. Moreover, exploring ensemble methods that combine the strengths of both Time Series Analysis and LSTM Neural Networks could potentially yield more robust predictions.

In conclusion, the comparative analysis highlights the potential of LSTM Neural Networks in stock price prediction, as indicated by their superior forecasting accuracy compared to Time Series Analysis using ARIMA. However, the dynamic nature of financial markets calls for ongoing research and refinement of these methodologies to enhance their robustness and practicality in real-world investment decision-making.

## 4. Conclusion

In this comparative study, two distinct methodologies, time series analysis using ARIMA and LSTM Neural Networks, were explored for stock price prediction. The analysis encompassed comprehensive dataset preprocessing, model development, and rigorous evaluation using consistent metrics. The results presented intriguing insights. LSTM Neural Networks consistently outperformed Time Series Analysis using ARIMA, exhibiting superior forecasting accuracy. The LSTM model's ability to capture intricate temporal dependencies showcased promise in modeling the complex and non-linear nature of financial markets. These findings underscore the adaptability of neural networks in addressing the inherent unpredictability of stock price data.

However, it is crucial to acknowledge the inherent challenges of stock price prediction, influenced by market sentiment, external events, and unforeseen shocks. Further research, parameter tuning, and adaptation to evolving market conditions are imperative for practical implementation. While LSTM Neural Networks demonstrated their prowess in this study, the quest for accurate stock price prediction remains an ongoing endeavor. Combining the strengths of both methodologies and exploring ensemble methods may unlock more robust and resilient forecasting solutions. This study serves as a steppingstone in the journey toward harnessing advanced techniques in financial forecasting, offering valuable insights for investors, analysts, and researchers navigating the complexities of the financial world.

## References

[1] Wang, Mengxian, et al. Stock price prediction for new energy vehicle enterprises: An integrated method based on time series and cloud models. Expert Systems With Applications, 2022.

[2] Gülmez, Burak. Stock price prediction with optimized deep LSTM network with artificial rabbit's optimization algorithm. Expert Systems With Applications, 2023.

[3] Shah, Jainil. Netflix Stock Price Prediction. Kaggle, 2022, www.kaggle.com/datasets/jainilcoder/netflix-stock-price-prediction.

[4] Hayes, Adam. Correlation: What It Means in Finance and the Formula for Calculating It. Investopedia, 2023.

[5] Theerthagiri, et al. Seasonal learning based Arima algorithm for prediction of Brent Oil Price Trends. Multimedia Tools and Applications, 2023, 82: 24485–24504.

[6] Ahalpara, et al. Characterizing and modelling cyclic behaviour in non-stationary time series through multi-resolution analysis. Pramana, 2008, 71: 459–485.

[7] Li, Susan. Time Series Analysis, Visualization & Forecasting with LSTM. Towards Data Science, 2019.

[8] Ahalpara, et al. Characterizing and modelling cyclic behaviour in non-stationary time series through multi-resolution analysis. Pramana, 2008, 71: 459–485.

[9] Bayram, et al. DA-LSTM: A dynamic drift-adaptive learning framework for interval load forecasting with LSTM networks. Engineering Applications of Artificial In-telligence, 2023.

[10] Terbuch, Anika. Quality monitoring in vibro ground improvement-A hybrid ma-chine learning approach. Geomechanics and Tunnelling, 2022, 15: 658-664.