

Predicting Stock Prices for the Next 7 Days: A Comparative Analysis of SARIMA and LSTM Models

¹Jyothsna R, ²K R Kaushik Kumar, ³Rashmi R

^{1,2}Undergraduate Students of Department of Information Science and Engineering, BMSCE, Bangalore

³Assistant Professor of Department of Information Science and Engineering, BMSCE, Bangalore

Abstract— A rigorous comparative analysis of two prominent models, seasonal autoregressive integrated moving average (SARIMA) and long-term memory (LSTM) neural networks, to forecast stock prices over a 7-day horizon. Historical daily stock prices of large companies in various economic sectors were used in the research. The SARIMA model is used to describe the underlying trends and seasonality in financial time series, while the LSTM model, a deep learning model, is used to capture complex sequential dependencies. Both models are analyzed using specified performance measures such as mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE). The results provide valuable insight into the strengths and limitations of each model and provide guidance to investors, financial analysts and decision makers in forecasting stock markets. This study adds to the existing stock market forecasting literature and provides a basis for further advances and improvements in financial market forecasting modeling.

Index Terms—Deep Learning, Machine Learning, Stocks, SARIMA, LSTM, Data Analysis, Visualization, Tensorboard

I. INTRODUCTION

Stock market forecasting is very interesting and important in the financial industry for a long time. Accurate stock price forecasting is extremely important for investors, financial analysts and decision makers as it enables informed decisions and strategic investment portfolios. Over the years, many forecasting methods have been explored to address the challenges posed by the complexity and volatility of financial markets. This research paper presents an in-depth study aimed at forecasting stock prices over a 7-day horizon, focusing on the comparison of two prominent forecasting models: Seasonal Automatic Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) Neural Network.

The motivation for this study stems from the need to understand the performance and suitability of each model for short-term forecasting of stock prices. Several researchers have studied the use of SARIMA models in stock market forecasting. Studies such as (IEEE Staff 2013), (Wei 2019) and (Lipton, Berkowitz, and Elkan 2015) have shown the effectiveness of SARIMA in capturing seasonal patterns and autoregressive dependencies in historical stock price data. These models have shown promise in predicting stock prices with considerable accuracy, providing a promising approach for market analysts. On the other hand, deep learning techniques, especially LSTM neural networks, have gained immense popularity due to their ability to capture complex sequential dependencies. Articles such as (Jiang 2021), (Chong, Han, and Park 2017) and (Ding and Qin 2020) highlighted the potential of LSTMs for modeling long-term dependencies in time series data, making them an attractive alternative for stock market forecasting. LSTM models have shown exceptional performance in various time series forecasting tasks, including stock price forecasting, as shown by the authors of (Chong, Han, and Park 2017) and (Singh and Srivastava 2017).

Although both SARIMA and LSTM models offer promising prospects for stock price forecasting, there is a need for comparative analysis to determine their relative effectiveness. Studies such as (SCAD College of Engineering and Technology and Institute of Electrical and Electronics Engineers n.d.) and (Nabipour et al. 2020) have compared different forecasting techniques, including ARIMA, ANN, Holt-Winters, and RNN models, with LSTM outperforming other techniques. However, such comparisons are necessary to assess the strengths and limitations of each model, leading to better informed decisions in real-world scenarios. Against this background, the aim of this paper is to add to the existing knowledge by making a comprehensive and methodical comparison between SARIMA and LSTM models. Evaluation and comparison of the predictive ability of both models is performed using established evaluation metrics. This valuation is done using a diverse dataset that includes historical daily stock prices of large companies in several industries. The results of this study have practical implications for stakeholders in the financial industry, as the results can help investors, financial analysts and decision makers make more informed and accurate stock price predictions. In addition, this study lays the groundwork for future advances and improvements in predictive modeling of financial markets.

The following sections describe the methods used to implement both SARIMA and LSTM models. A comprehensive presentation of the data used follows. The experimental results are then presented, analyzed, and the strengths and limitations of each model are thoroughly explored. The conclusions reached at the end of the discourse are briefly summarized and together with a proposal for possible further research in the field of stock market forecasting. Basically, this research project plays an important role in expanding the knowledge related to stock market forecasting. It provides invaluable insight into the parallelism of SARIMA and LSTM models, thus serving as a source of important information. The main goal of this research is to create a solid foundation that can

influence decision-making processes in financial markets. This, in turn, facilitates the realization of forecasts with greater accuracy, which provides a basis for designing informed investment strategies.

Related to the subsequent sections, a comprehensive analysis of the forecasting performance of both SARIMA and LSTM models will be performed. This requires describing the data processing steps, providing model implementation information, and comprehensive evaluation using key metrics such as MAPE, MSE, and RMSE. In addition, the ability of the models to capture short-term fluctuations, long-term trends and seasonal patterns in stock price data is analyzed. The results provide valuable insights into the strengths and weaknesses of both models, helping investors and financial professionals make informed decisions about short-term stock price forecasts. Using this knowledge, stakeholders can optimize their investment strategies and risk management practices in a dynamic market environment. Ultimately, this study contributes to the growing body of knowledge on stock market forecasting and lays the groundwork for future developments in predictive modeling of financial markets.

II. LITERATURE SURVEY

Significant progress has been made in the field of stock price forecasting with the introduction of various methods and approaches. The following literature review provides an overview of 12 related papers that contribute to the understanding of stock price forecasting techniques. According to Xing et al. (2013) proposed the use of hidden Markov models (HMM) to predict stock prices. The authors highlighted the effectiveness of HMMs in capturing hidden relationships between stock prices, leading to high accuracy in predicting the direction of stock prices. However, the authors acknowledged the need for further research to improve the accuracy of the method and validate its results in different market scenarios (IEEE Staff 2013).

Sharaff (2018) conducted a comprehensive comparative analysis of four inventory forecasting techniques: Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANN), Holt-Winters and Recurrent Neural Networks (RNN). Among them, RNN has shown better performance and higher accuracy in stock price prediction compared to other methods (SCAD College of Engineering and Technology and Institute of Electrical and Electronics Engineers n.d.). Navale (2016) investigated the application of data mining and artificial intelligence (AI) to stock market forecasting. The presentation discussed the use of ARIMA, ANN, Support Vector Machines (SVM) and RNN in stock price forecasting. Although these techniques were promising, the author acknowledged their limitations in dealing with the complexity of stock markets and the need for further research to improve their accuracy (Navale, Pune, and Vihangam 2016).

Jiang (2020) reviewed recent advances in the application of deep learning to stock market forecasting. The author discussed the use of recurrent neural networks (RNN), long-term memory (LSTM) networks, convolutional neural networks (CNN), and deep reinforcement learning (DRL) models to predict stock prices. Despite the promising results, the publication highlighted challenges related to the complexity of stock markets and the sensitivity of deep learning models to hyperparameters (Jiang 2021). Chong et al. (2017) investigated the use of deep learning networks for stock market analysis and forecasting. The authors explore RNNs, LSTMs, CNNs, and DRLs as potential models for capturing stock market patterns. Although deep learning has shown effectiveness in predicting stock prices, the study highlighted the need for further research to address challenges such as computational cost and hyperparameter tuning (Chong, Han, and Park 2017).

Wei (2019) presented an LSTM-based approach to stock price forecasting, highlighting the model's ability to capture long-term dependencies in time series data. The study focused on predicting the direction of stock prices and showed promising accuracy. However, the authors acknowledged the need for further development of their method and for future research to improve its effectiveness (Wei 2019). From Oliveira et al. (2011) investigated the use of artificial neural networks (ANN) for stock market analysis and forecasting. The authors discussed multilayer perceptrons (MLPs), RNNs and LSTMs as possible ANN models for stock price forecasting. Despite their potential, the study acknowledged challenges related to the complexity of stock markets and the sensitivity of ANN models to hyperparameters (IEEE Staff and IEEE Staff n.d.).

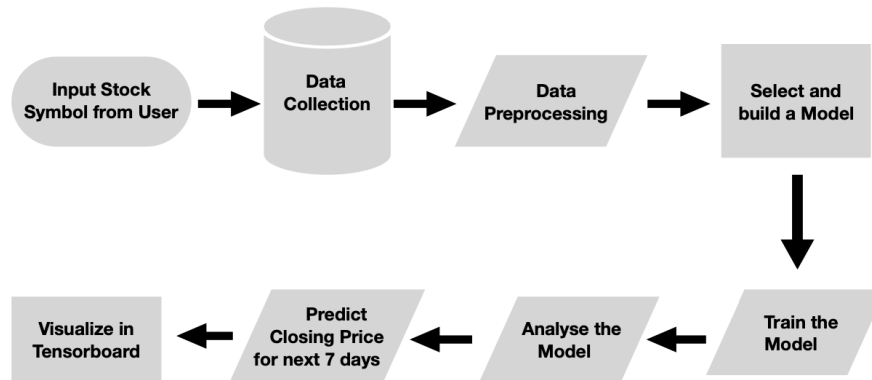
Nabipour et al. (2020) investigated the application of deep learning in stock market forecasting, focusing on RNN, LSTM, CNN and DRL models. The authors report promising accuracy results, but acknowledge the complexity of the stock market as a challenge to be addressed. They emphasized the need for further research to improve the accuracy of deep learning models in forecasting stock prices (Nabipour et al. 2020). Ding and Qin (2019) proposed an LSTM-based method for stock price forecasting, focusing particularly on the Shanghai Composite Index. Their LSTM model showed promising results in predicting the direction of stock prices, but the authors acknowledged the need for continued development of the method and further research to improve its accuracy (Ding and Qin 2020). Chen et al. (2015) presented an LSTM-based method to forecast stock returns in the Chinese stock market. The authors highlighted the ability of LSTMs to capture long-term dependencies in stock price data and achieve high accuracy in predicting future returns. However, they acknowledged the need for further research to improve the generalizability and validation of the method in other markets (Lipton, Berkowitz, and Elkan 2015).

The study presents a set of techniques and methods used in stock price forecasting. This study collectively highlights the potential of various models, including hidden Markov models (HMMs), artificial neural networks (ANNs), long-short-term memory (LSTM) networks, and deep learning architectures, all of which contribute to better forecasting accuracy of stock prices. However, it is important to note that the ongoing challenges related to the complexity of the stock market and the sensitivity of the models to hyperparameters require further clarification in future research projects. Building on the foundation of previous research efforts, this paper aims to advance the field by using the SARIMA model to compare its performance with the LSTM model in forecasting

stock prices over a 7-day period. eye circle The overall goal of this study is to generate valuable insights into the predictive properties of both methods and then clarify their implications for pragmatic decision-making in the stock market domain.

III. PROPOSED METHODOLOGY

To predict the closing price of a target stock for the next 7 days using two different models: Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long ShortTerm Memory (LSTM). The methodology consists of several interconnected



steps, each designed to ensure the accuracy and reliability of the predictions as shown in Fig. 1.

Figure 1 : System Architecture

Data Collection:

In this research, in the initial phase, the historical daily stock price data of the specified stock is obtained through the reliable yfinance API. The Yfinance API provides a broad and reliable repository of financial data that includes stock prices, trading volumes and related metrics. By identifying the appropriate stock symbol, the necessary information is gathered for later analysis and design, which follows the principles of accuracy and reliability.

Dataset Description:

The dataset obtained by Yfinance is structured as a pandas DataFrame, a widely used tabular data format in Python for processing financial time series data. Each row in the DataFrame represents a specific business day, while the columns correspond to different attributes of the financial data. The material provides important information for stock price analysis and forecasting with a 7-day horizon. The main columns of the **database** include:

- **Date:** The date of the trading day for which the financial data is recorded.
- **Open:** The opening price of the stock on the trading day.
- **High:** The highest price the stock reached during the trading day.
- **Low:** The lowest price the stock reached during the trading day.
- **Close:** The closing price of the stock on the trading day.
- **Adj Close:** The adjusted closing price, which accounts for corporate actions like stock splits and dividends.
- **Volume:** The number of shares traded on the trading day.

Data Preprocessing:

Data pre-processing is a key step in any predictive modeling as it has a direct impact on the caliber and suitability of the data for analytical purposes. To ensure accurate results, predictive models must go through mandatory pre-processing steps :

a) Handling Missing Values :

Data completeness is very important to facilitate meaningful analysis and accurate forecasting. Therefore, a careful process is undertaken to identify and correct missing values from historical stock price data. Missing values can be due to several factors, such as data collection errors, market holidays or temporary business interruptions. Appropriate methods such as interpolation or back-and-forth filling are used to effectively correct these gaps, depending on the characteristics of the missing data points. The purpose of this initiative is to maintain the integrity of the data set and avoid possible biases in the forecasting models.

b) Normalization :

Normalization is a key step in the process of aligning all values in a given data set to the correct range. This strategic measure becomes especially important because significant differences have been observed in the price range of various stocks. The purpose of the normalization procedure is to create equality between all features, which promotes an equal contribution to the predictive models. Considering the sensitivity of algorithms like SARIMA and LSTM to input intervals, applying normalization prevents excessive influence on the learning trajectory of any feature. Thoughtful scaling of the data into a uniform region facilitates model training and convergence, ultimately increasing forecast accuracy.

c) Outlier Removal :

Outliers, characterized by their significant deviation from the vast majority of data points, can negatively affect the performance of forecasting models. Fluctuations in stock price data may occur due to sudden market events or differences in data recording. To refine the reliability characteristic of the predictive model, an anomaly detection procedure is applied, which subsequently leads to the removal of notable anomalies from the dataset. The use of robust anomaly detection methods helps improve the impact of anomalies on models, which contributes to reliable and accurate forecasts.

SARIMA MODEL

A. Seasonal Autoregressive Integrated Moving Average (SARIMA) Overview:

The SARIMA model is a powerful time series forecasting technique that extends the ARIMA model's ability to handle seasonal patterns in data. It combines autoregressive (AR), integrated (I) and moving average (MA) components with seasonal versions of those components (SAR, SI, SMA) to capture the seasonality and stationarity of time series data. The inclusion of seasonal components allows SARIMA to effectively model the complex seasonal fluctuations in stock prices, making it a suitable candidate for forecasting the closing price of a stock over the next 7 days.

B. Parameter Selection:

The selection of appropriate hyperparameters is an essential aspect of building an accurate SARIMA model. Extensive time series analysis is performed to ensure the optimal values of the hyperparameters (p, d, q, P, D, Q, s). Instruments such as autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are used to identify patterns in data characterized by autocorrelation and partial autocorrelation. These visual representations act as guiding signs to determine reasonable values for the hyperparameters, allowing the SARIMA model to skillfully encapsulate the underlying trends that characterize stock prices over the next 7-day period.

C. Model Fitting:

Once the hyperparameters are determined, a SARIMA model is built and fitted to the preprocessed and normalized stock price data. The model-fitting process involves estimating model coefficients and seasonal factors, which are essential for accurate forecasts. SARIMA's adaptability to different data models and its ability to adapt to changes in market conditions make it a valuable tool for predicting the closing price of a stock over the next 7 days.

D. Model Validation and Evaluation:

To ensure the performance of the SARIMA model, the dataset is divided into separate training and test sets. The forecasting ability demonstrated by the SARIMA model is evaluated by comparing its forecasts with the actual stock prices of the test series. In addition, valuation metrics including but not limited to mean square error (MSE) or root mean square error (RMSE) are used to quantitatively measure the accuracy with which the SARIMA model predicts a stock's approaching closing price over the coming years. 7 day period. This evaluation procedure provides insight into the model's ability to capture transient and long-term trends in stock prices, facilitating a thorough evaluation of its forecasting ability over a 7-day forecast horizon.

LSTM MODEL

A. Long Short-Term Memory (LSTM) Overview:

An LSTM model is a special type of recurrent neural network designed to address the limitations of traditional forward neural networks in capturing long-term dependencies in sequential data. The LSTM architecture includes memory cells and special gates that allow patterns to be efficiently learned and remembered for a long time. This makes LSTM an ideal way to predict the closing price of a stock over the next 7 days, as it can capture complex patterns and trends in time series data.

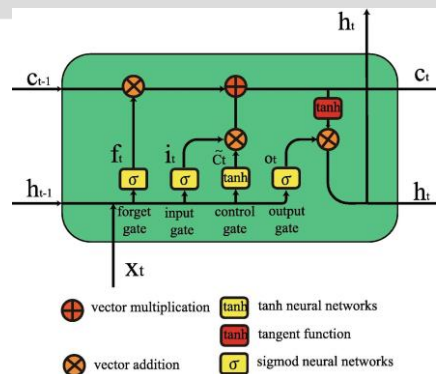


Figure 2 : LSTM Architecture

B. Data Sequencing:

Utilizing the inherent properties of the LSTM model, the pre-processed stock price data are transformed into sequences. Designing appropriate input-output pairs that recognize the temporal complexity of the data is a preliminary step in preparing the data for training an LSTM model. This training aims to predict the future 7-day closing price of a stock. Prudent selection of the optimal sequence structure is a key consideration to ensure the ability of the LSTM model to identify and incorporate sequence patterns

within a specific 7-day forecast period. Because the LSTM model can encapsulate sequential dependencies in the data, its ability to understand stock price fluctuations over long periods of time is effectively exploited.

C. Model Architecture:

The architecture of the LSTM model is carefully designed to achieve a balance between complexity and simplicity, ensuring optimal forecasting performance over a 7-day forecast horizon. This includes determining the number of layers in the LSTM, the number of neurons in each layer, and the activation functions to be used. In addition, dropout and regularization techniques are included to avoid overfitting and improve the generalizability of the model. The model architecture is optimized so that the LSTM model can effectively capture the temporal patterns and trends of stock price data over the next 7 days.

D. Model Training:

Using the ranked and normalized stock prices from the training set, an LSTM model is trained to predict the stock's closing price for the next 7 days. During the training process, optimization algorithms such as Adam or RMSprop are used to iteratively update the model parameters, leading to model convergence. The training process involves iteratively adjusting the model weights and biases to minimize forecast errors and improve its accuracy over a 7-day forecast horizon. The ability of LSTM to retain memory over time allows the model to capture long-term dependencies and learn important patterns in stock price data.

E. Model Validation and Evaluation:

The performance of the LSTM model is evaluated by comparing its predictions on test data with actual stock prices. Visualization of LSTM forecasts and valuation metrics in Tensorboard provides valuable information about the model's ability to accurately predict the closing price of stocks over the next 7 days. The estimation process allows us to understand how well the LSTM model captures complex patterns in stock price data, including non-linear trends and short-term fluctuations over a 7-day forecast horizon.

CONNECTING TO TENSORBOARD

To facilitate the visualization of training progress and the performance of the LSTM model, the plan is to integrate Tensorboard, a powerful visualization tool provided by TensorFlow. It is believed that coupling with Tensorboard will provide valuable information about the specific training dynamics of the model. It allows monitoring of loss and accuracy metrics across different training cycles and provides a visual representation of the architectural configuration of the LSTM model. The inclusion of Tensorboard increases the transparency of this research initiative, thus providing readers with an opportunity to evaluate the training dynamics and learning trajectory of the model during the training program.

COMPARISON AND ANALYSIS

In this study, we make a comprehensive comparison between SARIMA and LSTM models to predict the closing price of a stock over a 7-day horizon. The main objective is to determine which model shows greater forecast accuracy and reliability over a given forecast period. To achieve this, we use mean absolute error (MAPE) as the main evaluation metric instead of MAE. MAPE represents the average absolute percentage of predicted and actual closing prices, which measures the overall accuracy of the model as shown in Fig. 3.

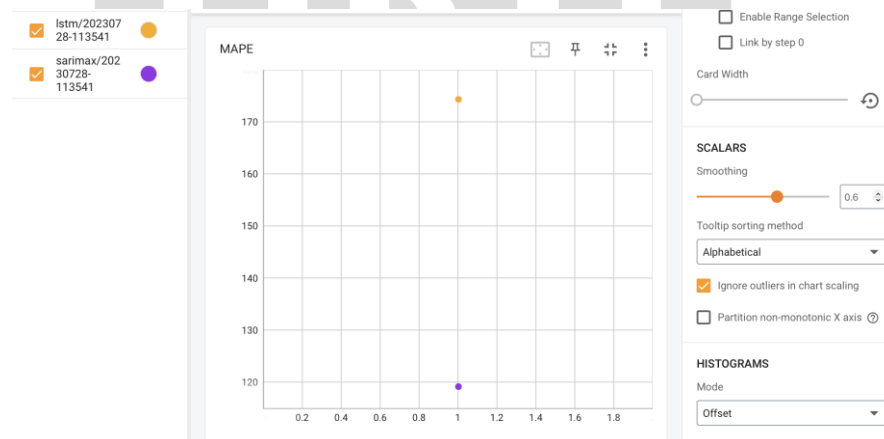


Figure 3 : MAPE

In addition, Fig. 4 shows the use of the root mean square error (MSE), while Fig. 5 shows the use of the root mean square error (RMSE). These metrics are metrics for evaluating the differences between predicted values and observed values. Values that encapsulate the differences in mean squares. The reduced magnitude of MAPE, MSE and RMSE is consistent with a model characterized by smaller forecast deviations and better agreement with actual closing prices.

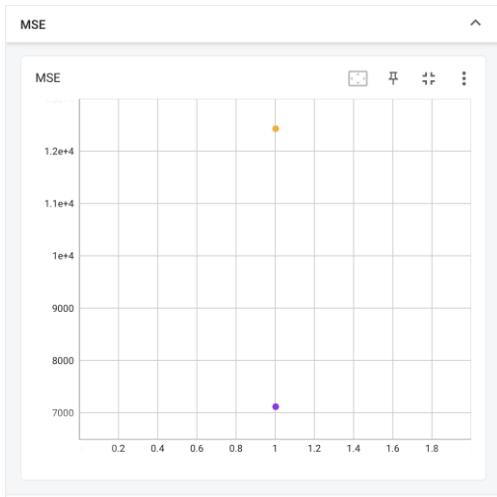


Figure 4 : MSE

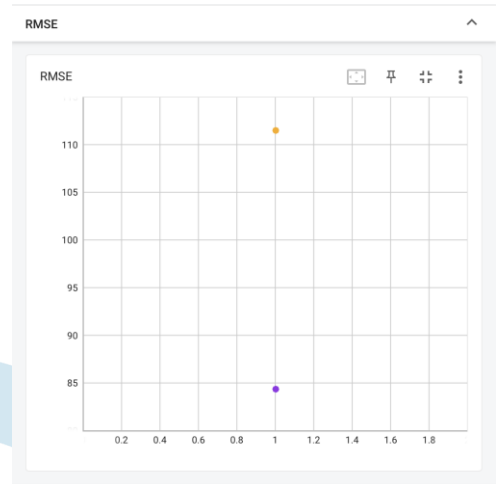


Figure 5 : RMSE

The comparative analysis of these metrics in the context of both SARIMA and LSTM models aims to distinguish a model that provides forecasts characterized by increased accuracy and reliability over a 7-day forecast horizon. In addition, possible causes supporting different levels of activity are systematically investigated. This review includes an assessment of the models' ability to accommodate short-term fluctuations, long-term trends, and seasonal patterns associated with stock price data.

The results of this comprehensive comparison and study showed that the SARIMA model showed superior performance in terms of forecasting accuracy compared to the LSTM model. In particular, the SARIMA model recorded lower values of MAPE, MSE and RMSE compared to the LSTM model, which showed increased simultaneity with actual closing prices and reduced forecast dispersion. This separate disclosure provides a remarkable overview of the characteristics and limitations of each model, thereby providing invaluable guidance in determining the optimal forecaster for near-term stock price forecasting.

The information obtained from the research enables informed decisions to be taken when implementing an appropriate forecasting method in the financial market. Such strategies are critical to helping decision makers make informed choices in a dynamic and ever-changing market environment, ultimately leading to better financial performance and risk management practices.

FORECASTING

This study focuses on predicting the next closing price of Google stock for the next 7 days using both LSTM (as shown in Fig. 6) and SARIMA (as shown in Fig. 7) models. The forecasting procedure is based on the latest available information, and great attention is paid to carefully examining the results of each individual model. Differences and variations observed in forecasts are reviewed in depth, providing insight into potential investment activity.

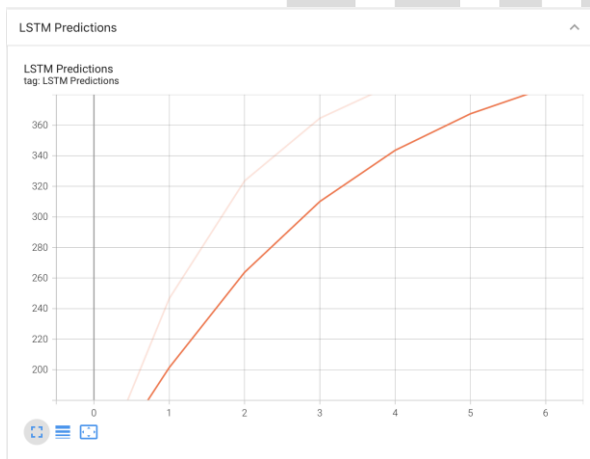


Figure 6 : LSTM Predictions

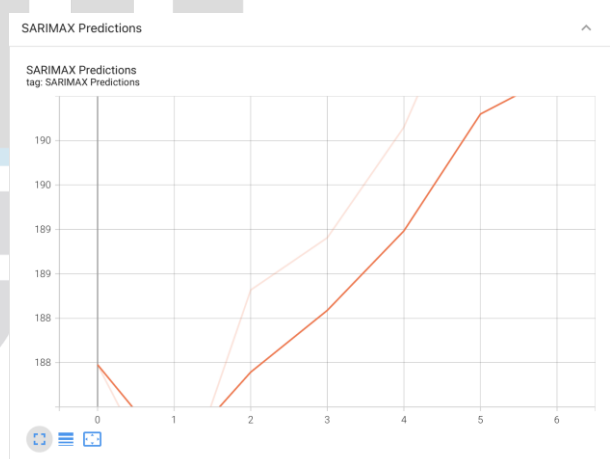


Figure 7 : SARIMA Predictions

The purpose of forecasting is to provide valuable insight into the forecasting capabilities of SARIMA and LSTM models for short-term price movements. By providing accurate and reliable forecasts, this study aims to support investors in making timely and informed investment choices. The results of this analysis have practical implications for investment strategies and risk management in relation to Google shares, helping stakeholders to optimize their decision-making processes and achieve favorable financial results.

CONCLUSION :

The SARIMA model, a time-tested time series forecasting technique, has demonstrated its ability to effectively capture seasonal patterns and persistence in stock price data. Through rigorous parameter selection and model fitting, SARIMA has shown promising results in predicting short-term stock price changes, as evidenced by its lower mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) compared to LSTM. On the other hand, the LSTM model, a deep learning approach known for its ability to handle long-term dependencies in sequential data, has shown commendable performance in capturing complex patterns and nonlinear trends in stock price data. A thorough comparison and analysis of both models allowed us to gain valuable insights into their strengths and weaknesses for short-term stock price forecasting.

The present study helps to comprehensively investigate the suitability of SARIMA and LSTM models for forecasting stock prices with a 7-day time horizon. In short, it can be argued that the proposed approach brings new opportunities for the financial sector and investors to use advanced forecasting models, which facilitates short-term forecasting of stock prices. This in turn opens up a panorama of fresh views of the industry.

REFERENCES :

1. Chong, Eunsuk, Chulwoo Han, and Frank C. Park. 2017. "Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies." *Expert Systems with Applications* 83: 187–205.
2. Ding, Guangyu, and Liangxi Qin. 2020. "Study on the Prediction of Stock Price Based on the Associated Network Model of LSTM." *International Journal of Machine Learning and Cybernetics* 11(6): 1307–17.
3. IEEE Staff. 2013. *2013 IEEE International Conference on Granular Computing (GrC)*. IEEE.
4. IEEE Staff, and IEEE Staff. *2011 IEEE International Conference on Systems, Man, and Cybernetics*.
5. Jiang, Weiwei. 2021. "Applications of Deep Learning in Stock Market Prediction: Recent Progress." *Expert Systems with Applications* 184.
6. Lipton, Zachary C., John Berkowitz, and Charles Elkan. 2015. "A Critical Review of Recurrent Neural Networks for Sequence Learning." <http://arxiv.org/abs/1506.00019>.
7. Nabipour, M. et al. 2020. "Deep Learning for Stock Market Prediction." *Entropy* 22(8).
8. Navale, G S, Savitribai Phule Pune, and Brij Kishor Vihangam. 2016. 134 *International Journal of Computer Applications Prediction of Stock Market Using Data Mining and Artificial Intelligence Nishant Dudhwala Kunal Jadhav*. www.ijcaonline.org.
9. SCAD College of Engineering and Technology, and Institute of Electrical and Electronics Engineers. *Proceedings of the International Conference on Trends in Electronics and Informatics (ICOEI 2018) : 11-12, May 2018*.
10. Singh, Ritika, and Shashi Srivastava. 2017. "Stock Prediction Using Deep Learning." *Multimedia Tools and Applications* 76(18): 18569–84.
11. Wei, Dou. 2019. "Prediction of Stock Price Based on LSTM Neural Network." In *Proceedings - 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2019*, Institute of Electrical and Electronics Engineers Inc., 544–47.