Journal of Computers, Mechanical and Management

Journal homepage: www.jcmm.co.in

Volume 2 Issue 1

Comprehensive Study of Machine Learning Algorithms for Stock Market Prediction During COVID-19

Urvashi Rahul Saxena*, Parth Sharma, and Gaurav Gupta

Department of Computer Science and Engineering, JSS Academy of Technical Education, Noida, Uttar Pradesh, India 201301

Abstract

The stock market has always drawn investors' attention. Stock trend forecasting tools are in high demand because they facilitate the direct acquisition of profits. The more precise the results, the greater the likelihood of generating a greater profit. Statistically, only 3% of the Indian population invests in stocks, and at the time of COVID-19, that number was even lower, as the stock market was not based on general patterns and equations but rather on the emotional quotient of the people. Such a circumstance increased the stock market's vulnerability. The pandemic factor, such as the case of COVID-19, influences the stock market's trends. Technical analysis of market trends is a technique for interpreting past and present prices to forecast likely future prices. Several deep learning and machine learning algorithms are utilized to generate stock market forecasts, wherein LSTM and ARIMA models have been proven to produce reasonably accurate results. The prior works focused on individual models and their components to provide forecasts. Thus, this paper aims to compare the two well-established models and provide investors with models that work well with data and have appropriate parameter values. The LSTM and ARIMA models are presented because they provide appropriate results using technical analysis of the data set, and the results are to be compared. The closing values from the historic stock prices of the top three Indian hospitality industries were used as the data set. The results show that the individual models work well when the data matches the model and the appropriate parameter is used.

Keywords: Machine Learning; Stock Predicts; COVID-19; Market Analysis; ARIMA; LSTM

1 Introduction

A stock market is the assembly of buyers and sellers of commodities which represents stock sold to investors through various platforms or the commodities privately traded (stocks of private companies) [1, 2]. The stock market is highly incoherent and non-static, so predicting a stock's future value is always difficult for the people trading or investing in the market. There have been 24 major market crashes, among which the most significant occurred in 2007 (Global financial crisis). Due to these variations, some people consider investing in the stock market equivalent to betting. Furthermore, one indeed faces a loss in investing in the stock market if one is not knowledgeable enough about it. This reason itself makes stock forecasting an important need. Indians began investing in stocks, which have outperformed all other asset classes by as much as 60% since 2001. However, due to stock market volatility, only 3% of the Indian population invests directly in stocks [3]. Multiple variables, such as news, social media data, fundamentals, company production, government bonds, historical price, and the country's economics, make it difficult to construct an accurate model. In addition, the COVID-19 pandemic severely impacted the factors above, negatively impacting the stock market, as decisions during the pandemic period largely depended on the public's emotional quotient [4]. Stock market forecasting has become increasingly popular due to its complexity.

*Corresponding author: urvashirahulsaxena@jssaten.ac.in

Received: 06 October 2022; **Accepted:** 02 January 2023; **Published:** 28 February 2023

DOI: 10.57159/gadl.jcmm.2.1.23021.

Collections from the stock market are typically dynamic, non-parametric, chaotic, and noisy. Consequently, the rate of change in the stock market is regarded as a random process. Technical analysis, time-series forecasting, machine learning and data mining, and modeling and predicting volatility of Stocks are among the most prominent techniques used to forecast stock market prices [5, 6]. From the list of several techniques, the autoregressive integrated moving average (ARIMA) method, a statistical approach and the deep learning method, namely long short-term memory (LSTM), have been proven to give the best results concerning market prediction [7]. Several researchers have worked in the area of stock prediction while proving the efficiency of the abovementioned models. Considering the application of the ARIMA model, Khanderwal and Mohanty presented [8] an in-depth method for developing it for stock price prediction and demonstrated experimentally its indisputable ability to predict stock costs using data from three national stock exchange (NSE) sectors over three years. Almasarweh and Wadi [9] utilized the ARIMA model to forecast the stock market for the Jordanian banking sector using daily data from 1993 to 2017. The results showed that the ARIMA model has significant results for short-term prediction and will thus be useful for investments. Wadi et al. [10] demonstrated that the ARIMA model has significant results for short-term prediction while predicting the stock market based on closing price data sets for the selected firms over 8 years. Wahyudi [11] used the ARIMA model to predict the volatility of Indonesian stock prices in his study, taking into account its capability, simplicity, and wide acceptability for predicting stock price volatility. Babu and Reddy [12] used the ARIMA model to examine the behavior of daily exchange rates of the Indian Rupee (INR) against various foreign currencies over 5 years. The results were compared with the results obtained using certain complex models. The proposed ARIMA model outperformed several complex nonlinear models and produced acceptable results.

Likewise, the LSTM model application also gained interest among researchers working in stock prediction. Hochreiter and Schmidhuber created LSTM, an RNN architecture that better stores and retrieves information and serves as the foundation for subsequent models, in 1997. Since then, it has been utilized in several prediction challenges, including stock market forecasts [13, 14]. Roondiwala et al. presented, modeled, and projected the NIFTY 50 stock returns using the LSTM technique and five years of historical data. The acquired findings were outstanding, with a prediction accuracy of approximately 98% [15]. Chen et al. modeled and forecasted China stock returns using LSTM, utilizing historical China stock market data turned into 30-day-long sequences with 10 learning features and 3-day earning rate labeling. The model was calibrated using 900000 sequences for training and 311361 sequences for testing. The LSTM model enhanced the accuracy of stock return forecasts by 13% compared to random prediction [16]. Liu et al. utilized the LSTM approach to filter, extract feature value, evaluate stock data, and establish the prediction model for the associated stock transaction. The experimental findings indicate that the created model's prediction accuracy was greater than 70% [17]. Mehtab et al. utilized LSTM regression models to anticipate the future NIFTY 50 open values utilizing four distinct models that varied in their architecture and input data format. The results demonstrated that the LSTM-based models produce accurate outcomes. Though several research documents deal with ARIMA and LSTM models for prediction problem-solving, they have mostly been used individually. Thus, the current work aims to develop the best model for predicting and analyzing stock market values while considering the datasets of the selected three top firms in the Indian hospitality industry. The paper presents the comparative results of the two majorly used stock forecasting models, LSTM and ARIMA, to assist investors and buyers in making more informed decisions.

2 Methods

2.1 ARIMA model

ARIMA modeling is fundamentally an exploratory data-oriented method that allows for the adaptability of a model based on the data's structure. With the aid of the autocorrelation function and partial autocorrelation function, the stochastic nature of a time series can be modeled approximatively; information such as trends, random variations, periodic components, cyclic patterns, and serial correlation can be determined. Consequently, it is possible to predict the future values of the series with a certain degree of precision [18]. Model identification, parameter estimation, and diagnostic testing are the steps in developing an ARIMA predictive model [19]. In the statistical literature, the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) technique is well-established [20]. It combines the differenced autoregressive and moving average models [21]. The model has proven to be one of the most prominent methods for financial forecasting and has demonstrated the ability to generate accurate short-term projections. The future value of a variable in the ARIMA model is a linear combination of past values and past errors [18, 20, 21], as expressed in Eq. [1].

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + +\varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_1 \varepsilon_{t-1} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{1}$$

where Y_t is the actual value, ε_t is the random error at t, φ_i and θ_j are the coefficients, and p and q are integers often referred to as autoregressive and moving averages, respectively. The Algorithm [1] isused for the ARIMA model in the current study:

2.2 LSTM model

Long short term memory (LSTM) is one of the most widely used deep learning models currently, especially for time series prediction, which is a difficult problem to solve due to long-term trends, seasonal and cyclical fluctuations, and random noise [22]. The ability of the LSTM to memorize data sequences distinguishes it from other RNNs [23]. LSTMs have longer memories and can learn from inputs separated by lengthy time lags.

Algorithm 1 ARIMA Model Algorithm

Require: Historical stock price dataset

Ensure: Stock price prediction based on stock price analysis

1: Import/load libraries (NumPy, Pandas (lag_plot), matplotlib, Stats models (ARIMA))
2: Read NSE- Hotel Leela Venture Limited/InterContinental Hotels Group (IHG)/Marriott International Incorporation (MII) dataset.
3: Plot the autocorrelation graph with a lag of 3 to check if there is any correlation in the data.
4: Plot the historic data in the date-time format.
5: Split the data into two variables, train_data (80%) and test data (20%).
6: Build the ARIMA model with the parameters (4, 1, 0).
7: Fit the model using the training dataset, generate prediction for each element on the test dataset.
8: Evaluate the model using RMSE (root mean squared error) loss function.
9: Plot the final output as the predicted price.

An LSTM consists of three gates: an input gate that determines whether to accept a new input, a forget gate that deletes irrelevant information and an output gate that determines what information to output. These three gates are analog gates based on the sigmoid function that operates between 0 and 1 [22]. Figure 1(a) represents the simplest form of a recurrent neural network (RNN), which considers not only its current input but also the output of the preceding RNNs. Figure 1(b) represents the LSTM cell, a special type of RNN capable of handling long-term dependencies compared to simple RNN. Figure 1(c) represents three interconnected cells of LSTM (used in the present work). Figure 1(d) represents the exploded view of the single LSTM cell. The Algorithm [2]is used for the LSTM model in the current study

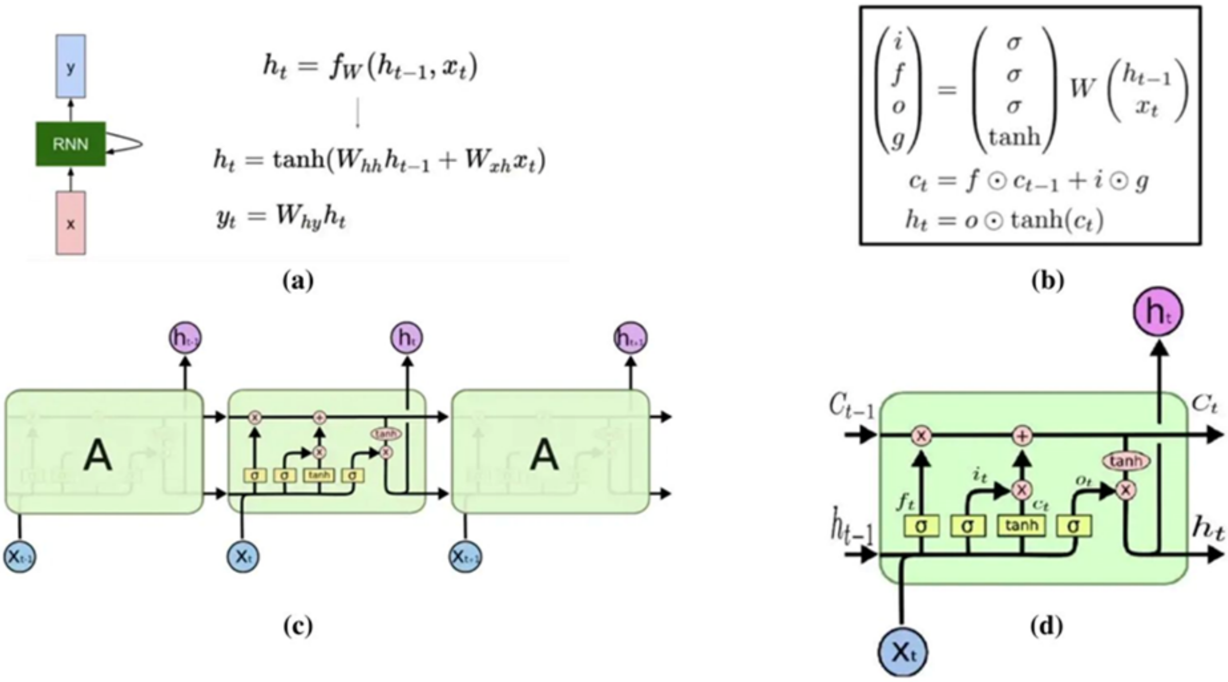

Figure 1: **(a)** Simple RNN cell representation; **(b)** LSTM cell representation; **(c)** Visual representation of interconnected LSTM cells; **(d)** Visual representation of single LSTM cell.

2.3 Data collection

The literature showed that during COVID-19, the hospitality sector faced the worst and steep downfall. Moreover, the current study is focused on the hospitality sector of India. Thus, the data focused are from Hotel Leela Venture (HLV) Limited, InterContinental Hotels Group (IHG) and Marriott International Incorporation (MII). Generally, two sources publish companies' stock prices in India: the Bombay stock exchange (BSE) and the National stock exchange (NSE). Also, the literature shows that closing price forecasting is an important rule in finance and economics, which has prompted researchers to develop a fit model for forecasting accuracy [10]. Thus, in the present study, the data concerning the closing stock prices of the selected firms is collected from NSE. Data were directly downloaded using YAHOO FINANCE and filtered by the date.

Algorithm 2 LSTM Model Algorithm

Require: Historical stock price dataset
Ensure: Stock price prediction based on stock price analysis

1: Import/load libraries (NumPy, Pandas, matplotlib, Keras models (Sequential), Keras layers (LSTM, Dropout, Dense), Scikit learn).
2: Read NSE- Hotel Leela Venture Limited/InterContinental Hotels Group (IHG)/Marriott International Incorporation (MII) dataset.
3: Plot the historic data of the dataset in date-time format.
4: Sort the data in ascending order according to the date.
5: Scale the data using MinMaxScaler, split the data into two variables, train_data (90%) and valid_data (20%).
6: Create training dataset, convert them into a Numpy array, reshape into 3-dimensional data.
7: Create testing dataset, convert them into a Numpy array, reshape them into 3-dimensional dataset, undo scaling of the model.
8: Build and compile the model using the training data set.
9: Use the output of the last layer as a prediction of the next time step.
10: Repeat steps 8 and 9 until optimal results are achieved.
11: Plot the final output as the predicted price.
12: Evaluate the accuracy using the RMSE (root mean squared error) function.

2.4 Data analysis

Initially, the prediction results obtained using two technologies: Random forecasting and the ARIMA model, were compared (during the pilot study), wherein the ARIMA model proved to be efficient. Moreover, considering the complexity of financial time series, combining deep learning with the concept of financial market prediction is regarded as one of the most efficient ways. Thus, an algorithm is proposed for predicting future values wherein the special type of RNN, namely LSTM, is implemented. The results obtained using ARIMA are compared with those obtained using the proposed LSTM model, and the same has been discussed in the present article. Values from the beginning in the first sliding window are used to predict the price in the following window.

3 Results and Discussion

For the implementation of the discussed ARIMA model, the closing time series data were first decomposed into three components: trend, seasonality and noise. The two-year ARIMA result concerning the stationary graph for HLV, one of the selected firms, is represented in Figure 2.

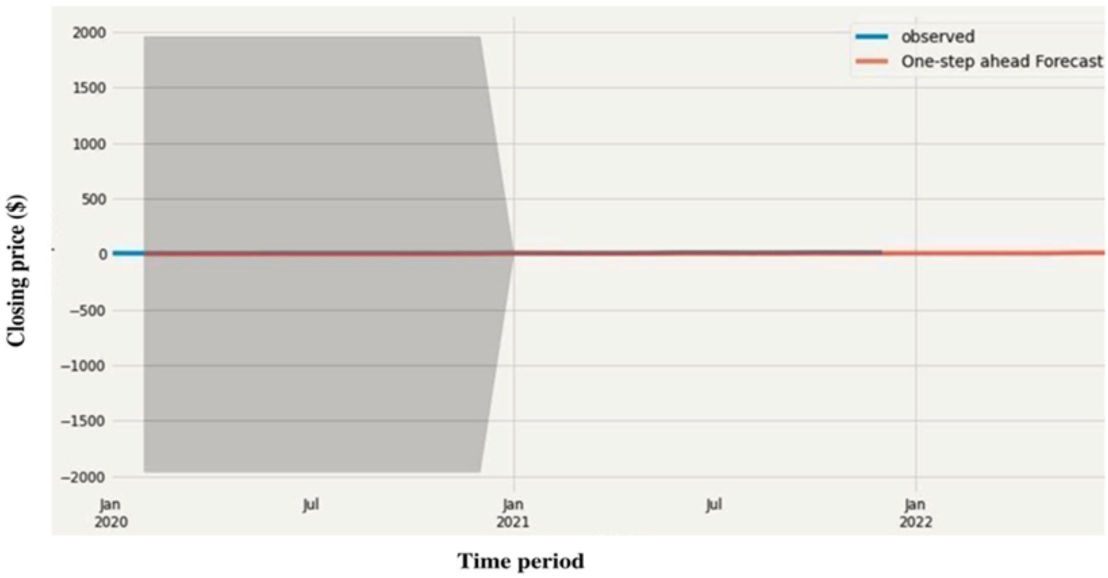

Figure 2: ARIMA results for one of the selected firms.

For LSTM, the data obtained were initially preprocessed, wherein only the stock's closing value was taken and mapped with the indexes. The values were transformed in a range between 0 and 1, using the minmax scalar, as the LSTM is known to be sensitive to the data scale. After the accomplishment of the transformation, the data was split into test and train. The split data was converted to a dataset matrix per the LSTM model requirements. The root mean square error (RMSE) values were calculated, and the predicted future values were plotted, as shown in Figure 3. The accuracy of the calculated value was determined as, Accuracy $= 100 - \text{RMSE}$.

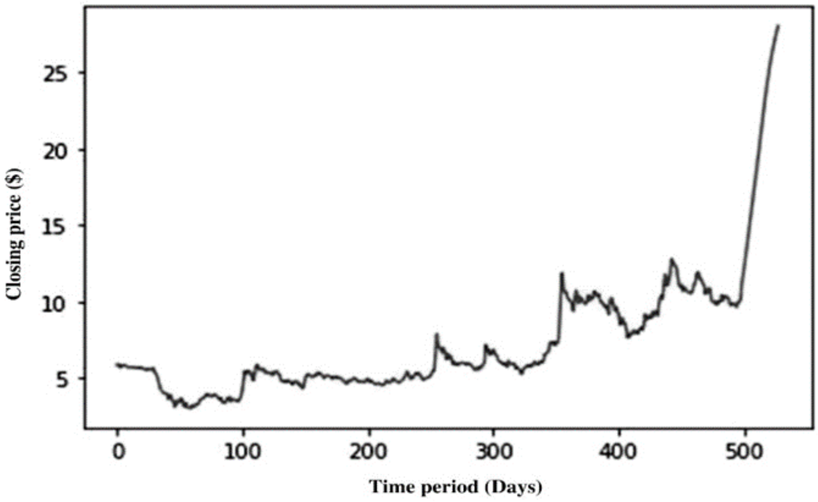

Figure 3: LSTM results for one of the selected firms.

The consolidated results concerning the RSME values obtained using ARIMA and LSTM models for the selected data set of the hospitality industry are given in Table 1. The same has been depicted graphically in Figure 4. From the results; it is evident that the results obtained using the LSTM model outperform the results obtained using the ARIMA model. The average accuracy level of values obtained using the LSTM model is 97.31%, and that obtained using the ARIMA model is 82.73%. The outstanding effectiveness of the RNN machine learning-based LSTM method to the chosen stock market prediction issue is attributable to the ”iterative” optimization technique employed by these approaches to get optimal outcomes. Iterative refers to repeatedly obtaining results and selecting the most ideal one, i.e., the iteration that minimizes errors. As a result, iterations aid in the transformation of an under-fitting model into a model that is ideally suited to the data.

Table 1: Closing price and prediction accuracy comparison of LSTM and ARIMA results.

Hospitality sector	**Closing values**		**Actual**	**Accuracy**	
	LSTM	**ARIMA**		**LSTM**	**ARIMA**
HLV Limited	5.11	4.40	5.3	96.42%	83.02%
Intercontinental Hotels Group	58.88	48.46	58.95	99.88%	82.21%
Marriot International Inc.	114.07	98.93	119.28	95.63%	82.94%

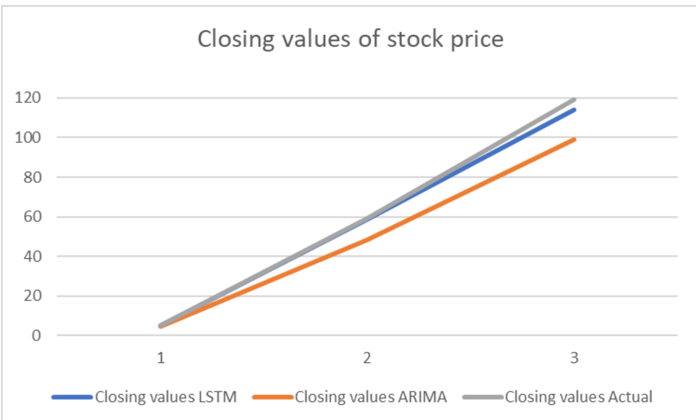

Figure 4: Comparative analysis of RSME values obtained using the ARIMA and LSTM models.

Since ML methods such as LSTM employ gradient descent to enhance their models, it makes sense to put the whole dataset through a single network numerous times to update the weights and achieve a more accurate prediction model. Another possible cause might be the trained model’s rolling characteristic. Since a model is refined in a rolling scenario in each round, a whole new LSTM model is being trained, and the weights are updated for each new model. The ARIMA model, though, showed a lower accuracy; it was very well above the acceptable value of 75%. An observation is that the time taken to process the ARIMA model was more. Thus, if faster processing devices are used or a data set with fewer empty values is used, the ARIMA model can produce better results.

4 Conclusion

The impact of COVID-19 on the global hospitality industry was significant. These industries suffered enormous losses during the pandemic, and the irregularity also affected other industries. The objective of this paper was to develop models that can provide investors with an overview of the market's general tendencies. Three datasets were used to construct and evaluate the model for NSE-Hotel Leela Venture (HLV) Limited, InterContinental Hotels Group (IHG), and Marriott International Incorporation (MII), with 80% of the data used for training and 20% for testing. The accuracy achieved was acceptable (nearly greater than 75%). Two models: ARIMA and LSTM, were used. The historical stock data for any given company is extremely extensive. RNNs, particularly LSTM models, have memory cells that only retain relevant information for a specific step and thus produce accurate results. In this regard, LSTM outperformed the ARIMA method because it only considers the most important variables for forecasting. During the analysis, it was discovered that the processing time required for ARIMA is significantly longer. However, if faster processing devices are used or a data set with fewer empty values is used, the ARIMA model can produce better results. The LSTM model made accurate predictions but tended to fail when any attribute values were missing. In simple words, it could be concluded from the work that the LSTM provided a better result than the ARIMA model, but then both models produce reasonably accurate results and could be a great asset for stock traders.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research did not receive any grants from governmental, private, or nonprofit funding bodies.

Author Contribution

Urvashi Rahul Saxena: Supervision, Writing- Original draft preparation, Writing- Reviewing; Parth Sharma: Conceptualization, Visualization, Investigation, Methodology, Data curation, Writing- Reviewing; **Gaurav Gupta**: Conceptualization, Visualization, Investigation, Methodology, Data curation, Writing- Reviewing.

References

[1] O. Bustos and A. Pomares-Quimbaya, ”Stock market movement forecast: A Systematic review,” Expert Systems with Applications, 156, p. 113464, 2020.

[2] D. Shah, H. Isah, and F. Zulkernine, ”Stock market analysis: A review and taxonomy of prediction techniques,” International Journal of Financial Studies, 7 (2), p. 26, 2019.

[3] K. Chitra and V. R. Sreedevi, ”Does personality traits influence the choice of investment?”, IUP Journal of Behavioral Finance, 8 (2), pp. 47–57, 2011.

[4] A. Nayak, M. M. M. Pai, and R. M. Pai, ”Prediction models for indian stock market,” Procedia Computer Science, 89, pp. 441–449, 2016.

[5] A. Subasi, F. Amir, K. Bagedo, A. Shams, and A. Sarirete, ”Stock market prediction using machine learning,” Procedia Computer Science, 194, pp. 173–179, 2021.

[6] D. P. Gandhmal and K. Kumar, ”Systematic analysis and review of stock market prediction techniques,” Computer Science Review, 34, p. 100190, 2019.

[7] Q. Ma, ”Comparison of ARIMA, ANN and LSTM for stock price prediction,” E3S Web of Conferences, 218, p. 01026, 2020.

[8] S. Khandelwal and D. Mohanty, ”Stock price prediction using ARIMA model,” International Journal of Marketing & Human Resource Research, 2 (2), pp. 2746–4040, 2021.

[9] M. Almasarweh and S. AL Wadi, ”ARIMA model in predicting banking stock market data,” Modern Applied Science, 12 (11), p. 309, 2018.

[10] S. AL Wadi, M. Almasarweh, and A. A. Alsaraireh, ”Predicting closed price time series data using ARIMA model,” Modern Applied Science, 12 (11), p. 181, 2018.

[11] S. T. Wahyudi, ”The ARIMA model for the Indonesia stock price,” International Journal of Economics and Management, 11 (SpecialIssue1), pp. 223–236, 2017.

[12] B. A. Reddy SK, ”Exchange rate forecasting using ARIMA, neural network and fuzzy neuron,” Journal of Stock & Forex Trading, 04 (03), 2015.

[13] Z. Jin, Y. Yang, and Y. Liu, ”Stock closing price prediction based on sentiment analysis and LSTM,” Neural Computing and Applications, 32 (13), pp. 9713–9729, 2020.

[14] S. Hochreiter and J. Schmidhuber, ”Long short-term memory,” Neural Computation, 9 (8), pp. 1735–1780, 1997.

[15] M. Roondiwala, H. Patel, and S. Varma, ”Predicting stock prices using LSTM,” International Journal of Science and Research, 6 (4), pp. 1754–1756, 2017.

[16] K. Chen, Y. Zhou, and F. Dai, ”A LSTM-based method for stock returns prediction: A case study of China stock market,” in Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015, Oct. 2015, pp. 2823–2824.

[17] S. Liu, G. Liao, and Y. Ding, ”Stock transaction prediction modeling and analysis based on LSTM,” in Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications, ICIEA 2018, May 2018, pp. 2787–2790.

[18] S. L. Ho and M. Xie, ”The use of ARIMA models for reliability forecasting and analysis,” Computers & Industrial Engineering, 35 (1–2), pp. 213–216, 1998.

[19] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, ”Using multivariate statistics,” 5th ed., Boston, MA: Pearson, 2007.

[20] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, ”Stock price prediction using the ARIMA model,” in Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014, Mar. 2014, pp. 106–112.

[21] V. Kotu and B. Deshpande, ”Time series forecasting,” in Data Science, Elsevier, 2019, pp. 395–445.

[22] A. Yadav, C. K. Jha, and A. Sharan, ”Optimizing LSTM for time series prediction in Indian stock market,” Procedia Computer Science, 167, pp. 2091–2100, 2020.

[23] J. Shah, D. Vaidya, and M. Shah, ”A comprehensive review on multiple hybrid deep learning approaches for stock prediction,” Intelligent Systems with Applications, 16, 2022.