A COMPARATIVE STUDY OF STOCK PRICE PREDICTION USING ARIMA - LSTM TIME SERIES MODELS

Abstract

Time series prediction is incredibly important and difficult; external factors can also affect prediction results. Therefore, this study compares the following two models to see which makes better predictions. The traditional model is the "Autoregressive Integrated Moving Average" (ARIMA). Over time, new variables like SRIMA (seasonal ARIMA) have emerged. This model is better for short-term prediction than long-term prediction. Another model, CNN and RNN, is more dependent on artificial intelligence data analysis and closer to upcoming technologies. This research focuses on LSTM neural networks, which can learn from historical data and relate to present data. According to this paper's data, LSTM's data prediction is superior to the ARIMA model.

Keywords: LSTM, ARIMA, Prediction, D L, ML.

Authors

Pappula Ashok
Research Scholar
Department of Mathematics
GITAM Hyderabad
Telangana,India.

Dr D. Mallikarjuna Reddy
Professor
Department of Mathematics
GITAM Hyderabad
Telangana, India.

I. INTRODUCTION

As of March 31, 2023, State Bank of India's market worth was Rs. 5,954,418.30 Cr. Bank started in 1806. The bank's NSE code is SBIN, and the BSE code is 500112. SBI, formerly the Imperial Bank of India, is India's largest bank. World's 43rd largest bank. This public bank owns SBI Life Insurance Ltd. and SBI General Insurance. The largest public sector commercial bank is State Bank of India. For more than 20 years, it has served consumers across India. The Reserve Bank of India, India's government bank, purchased 60% of Imperial Bank of India's stock in 1955 and renamed it State Bank of India.

Predicting the stock price of SBI BANK has become crucial, and it's important to utilise the right time series method to predict it. If stock prices can be reliably forecasted, investors will have access to vital decision-making information to help them decide how to invest in technology development. Stock price changes explain U.S. investment, especially for long-term samples and when cash flow variables are present. [3]For investors, equities make sense. Stocks may yield high returns. Stockholders receive dividends for their contributions to a company's growth. They can speculate by buying stocks. Buy low and sell high to make a profit. According to Lee et al.'s [4] research, a market prediction is crucial to a successful investment plan. By understanding the market, this can be done. SBI BANK's closing price is predicted using numerous machine learning methods in this study. One thousand two hundred ninety-eight historical data points from January 1, 2018, to March 31, 2023, were used to forecast SBI BANK's closing price. ARIMA and LSTM models are used for predictions these. The model that best predicts SBI BANK's closing price can be determined by comparing its training and testing results.

II. LITERATURE REVIEW

A key part of time series data analysis is making predictions. How time series data is analysed, and predictions are made relies on the data type and situation. Seasonality, economic shocks, unexpected events, and internal changes can also change the estimate [5]. This study uses two-time series prediction models to compare how well they predict SBI BANK stock data and how often they get it wrong. This lets the best model be picked. This study looks at stock info from SBI BANK for many reasons: First, SBI BANK has a good reputation worldwide, which means that its facts are reliable and accurate. Second, the way its data changes shows how the economy changes at different times, which can be used to predict how the economy will change.

ARIMA starts this investigation [6]. It's classical. Method steps: Moving average predicts after model fitting. This linear regression-based method developed SARIMA and ARIMAX. These models work short-term but fail long-term [7]. Machine learning, specifically deep learning, uses artificial intelligence data analysis to eliminate data analysis, making the model data-driven rather than model-driven. Training the best application field learning model is achievable. CNNs are good for image recognition and other challenges, while RNNs are superior for modelling time series data and analysis [8]. Several RNN-based models exist. Most RNNs remember input data.

The LSTM model is additionally written about on this page. It acts like I/O. A feedback-based RNN-based model like LSTM can learn from past data and use gates like a grid structure to build a model of past and current data to go through the input data only once [9]. Most experiments show that the deep learning-based model performs better at making predictions about time series, especially about the future. In this study, the LSTM model does better than the ARIMA model [10].

III. METHODOLOGY

1. ARIMA

Time series data is predicted using the linear analysis ARIMA (p, d, q) model. The AR term's order is p. It has three terms: p, d, and q. A simple AR order p, AR(p), is a linear process:

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t \qquad \text{(I)}$$

where x_t is the stationary variable, c is a constant, the terms in i are autocorrelation coefficients at lags 1, 2, p, and the residuals, t, is Gaussian white noise series with mean zero and variance 2. When a non-stationary series becomes stationary, the MA term's order is q, and the differencing order is d. An MA order q can be formulated below:

$$x_t = \mu + \sum_{i=0}^{q} \phi_i \epsilon_{t-i} \qquad \text{(II)}$$

where μ is the expectation of x_t (usually zero), i terms are weights assigned to a stochastic term's current and prior values in the time series and 0-1. Suppose t is a zero-mean, 2-variance Gaussian white noise sequence. These two models form an ARMA order (p, q):

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t + \sum_{i=0}^{q} \phi_i \epsilon_{t-i} + \epsilon_t \qquad \text{(III)}$$

ACF behaviour and PACF plots are used to find models. After the first non-seasonal difference, the training data's ACF and PACF plots are shown below. The ACF and PACF plots are used to build random walk models with drift. White noise affects both graphs. Thus, speculating with a one-step out-of-sample projection is best. Multi-step out-of-sample forecasts with re-estimates are used here. The best estimate model is built after each re-fitting. The approach produces a prediction model and outputs the best guess's RMSE. Steps establish two data structures. These data structures, "prediction" and "history," store the constantly predicted values for the test data sets and the continuously added training data sets.

2. LSTM: The LSTM can deal with short-term and long-term trend reversals when projecting stock prices using indicators like MACD and RSI. Market abnormalities and pricing trends can be distinguished this way. Long-term observations can be stored and learned by the LSTM RNN model. A multi-step univariate prediction method was used to develop the algorithm. Mathematically, it's:

$$i_t = (W_i h_{t-1} + U_i x_t + b_i) \quad \text{(IV)}$$

$$f_t = (W_f h_{t-1} + U_f x_t + b_f) \quad \text{(V)}$$

$$o_t = (W_o h_{t-1} + U_o x_t + b_o) \quad \text{(VI)}$$

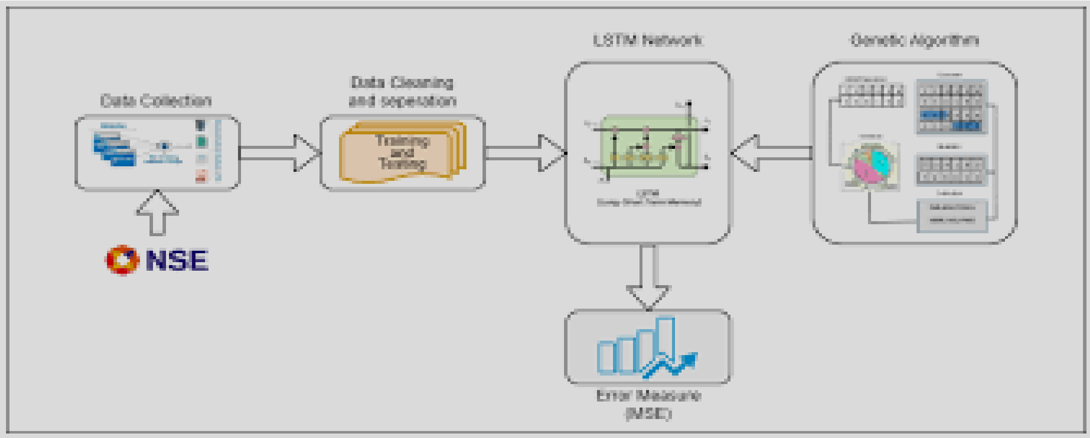

Figure 1: The LSTM Principle.

Where W , W_f , W_o, and W_c are the weights of h_{t-1}, respectively. The weights of x_t are denoted by the letters U , U_f, U_o, and U_c. b , b_f, b_o, and b_c are all examples of bias factors. The sigmoid function, denoted by , is responsible for ensuring that the values of i_t, f_t, and o_t, fall within the range of 0 to 1. The activation function is denoted by “real”. It is possible to demonstrate that the information stored in the concealed state at the preceding instant as well as the current state of the cell, C_t, both affect the magnitude of the quantity h_{t-1}.

Additionally, W_c , The gradient vanished for a reason that doesn't affect the cell state estimate. Adding a gating method to the training process solves the gradient disappearance problem and makes model predictions more accurate. The LSTM model has three logical control units: an input gate, an output gate, and a forget gate. This lets the RNN use long-term time information and fixes its short-term memory problem. They also have properties that make them multiply. By changing the weight at the edge of the neural network's memory unit, you can control how information flows and what state the memory cell is in. During the LSTM's learning process, data characteristics from time t are sent to the input layer. The activation function handles the outputs. Input the output result, the hidden layer output at time t-1, and cell unit data at time t-1 into the LSTM structure's node. Send the data to the next secret layer or output layer using the Input Gate, Transmit Gate, Forget Gate, and cell unit. LSTM structure node results are sent to output layer neurons, backpropagation error is calculated, and weights are changed. The LSTM model can be seen in Figure 1.

IV. RESULTS

From January 1, 2018, to March 31, 2023, Yahoo Finance provided SBI BANK stock data. The data set has parameters like Date, Open, Low, High, Adjusted Close, and Close. The ARIMA and LSTM models contain only the "Close" measure. The data set collection was divided into two parts. The first part, from January 1, 2018, to December 2022, is training data; the second part, from January 1, 2018, to March 31, 2023, is testing data. The last 62 observations were used to assess the model's performance, while the rest of the dataset was used to train it.

MAPE, RMSE, MAE, and MSE are used to assess each model's ability to predict SBI BANK's closing price. These find the residuals, or differences, between measured and predicted values. The measure compares model predictions for the same data collection. The indicator equations are:

$$\boldsymbol{MAPE} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \left| \frac{\mathbf{y_i} - \hat{\mathbf{y}}_\mathbf{i}}{\mathbf{y_i}} \right| \qquad \boldsymbol{RMSE} = \sqrt{\frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\mathbf{y_i} - \hat{\mathbf{y}}_\mathbf{i})^2}{\mathbf{N}}}$$

$$\boldsymbol{MAE} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} |\mathbf{y_i} - \hat{\mathbf{y}}_\mathbf{i}| \qquad \boldsymbol{MSE} = \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\mathbf{y_i} - \hat{\mathbf{y}}_\mathbf{i})^2}{\mathbf{N}}$$

Where Yi is the actual observations, $\hat{\mathbf{y}}_\mathbf{i}$ is the proposed model's predicted values, and N is the number of forecast values. The model with the lowest MAPE, RMSE, MAE, and MSE will be the best forecaster.

To begin, we perform the stationarity test on the ARIMA model using the ADF test. To begin, it is necessary to perform a regression analysis on the equation and examine the results of the ADF test to see whether or not it is possible to refute the hypothesis that H_0 is true. The p-value for the ADF test comes in at 0.959, and the result is a value of -0.94 for the test. This is a poor argument against the null hypothesis (H0). Thus we will accept the H_0 that the data are not stable. The first difference in the price of SBI Bank stock is used to transform the non-stationary data into stationary data to analyse the non-stationary data. This is done so that the non-stationary data can be dealt with. After accounting for the initial difference, the ADF test yields a result of -5.31, while the p-value sits at 9.12.

There is much evidence against the H_0, so we don't think the data are stable, which is what H0 says. The parameter d is the order of the frequency change when going from a time series that isn't stationary to one that is. Then, we look at how the ACF and PACF plots move to find models. After that, we can figure out the initial numbers for the ARIMA model's parameters: (0, 1, 0) for the AR order p, the order of differencing d, and the MA order q. The Error Matrix founding is shown in Table I. Based on the data, the average MAPE, RMSE, MAE, and MSE for the ARIMA and LSTM models are, respectively. This shows that LSTM makes a big difference in lowering the number of Errors. Error Matrix numbers show that LSTM models do better than ARIMA models.

In Figure 2, we have shown the LSTM model summary, including no hidden layers, batch size and dropout. In Figure 3, the SBI BANK stock price data set is divided into a Train and a Test plot. Figure 4(a) shows the predicted price of SBI BANK stock based on the ARIMA model Predictions and Test data. Figure 4(b) shows the closing prices that SBI BANK predicted and the actual closing prices tested. Figure 4(c) shows that, based on the SBI BANK stock price, the LSTM model is more accurate at predicting the future than the ARIMA model. By looking at the results, we can also see that the ARIMA model's forecast is directional, which comes from its model assumption liner, while the LSTM model is better at showing how the stock moves up and down.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 1, 64)             16896

 dropout (Dropout)           (None, 1, 64)             0

 lstm_1 (LSTM)               (None, 32)                12416

 dense (Dense)               (None, 1)                 33

=================================================================
Total params: 29,345
Trainable params: 29,345
Non-trainable params: 0
_________________________________________________________________
```

Figure 2: The Summary of the LSTM Model.

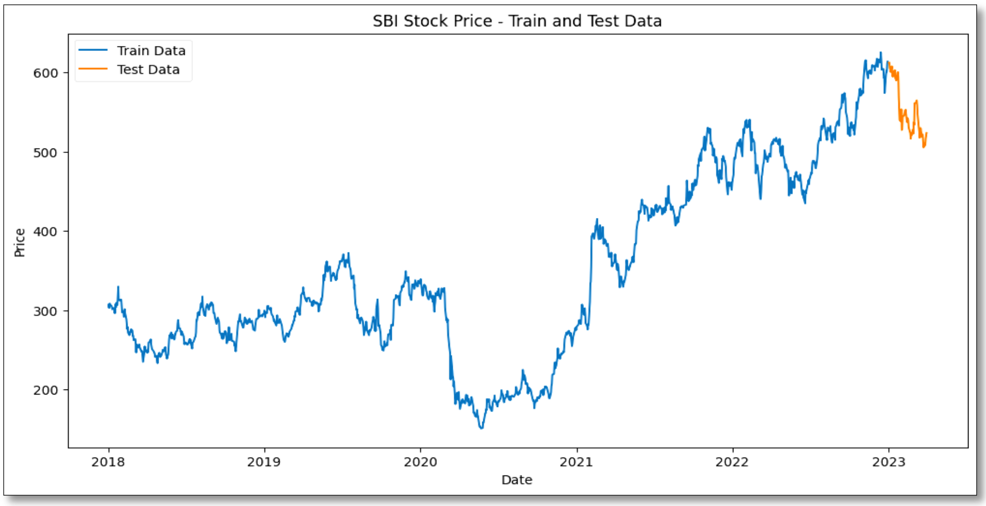

Figure 3: SBI BANK stock prices Train and Test Split Graph

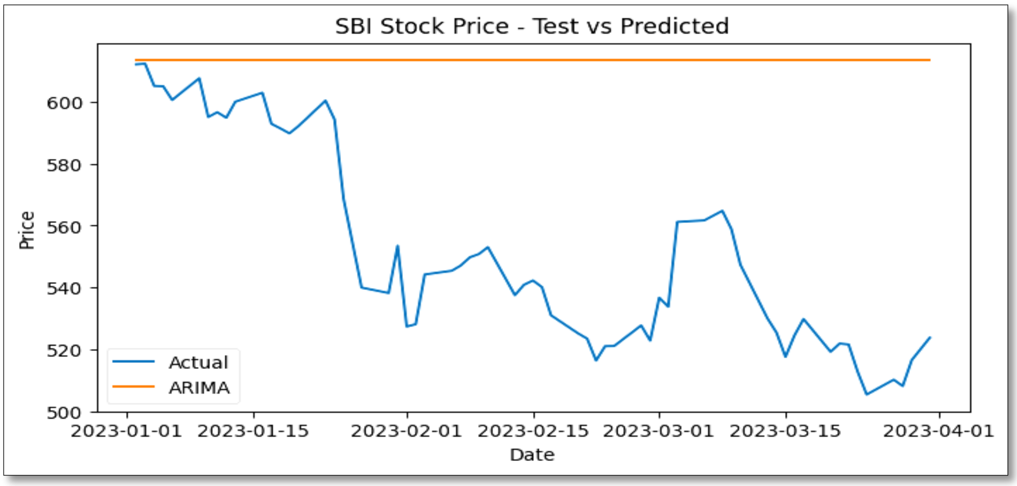

Figure 4(a): SBI BANK stock prices ARIMA Predictions and Test graph

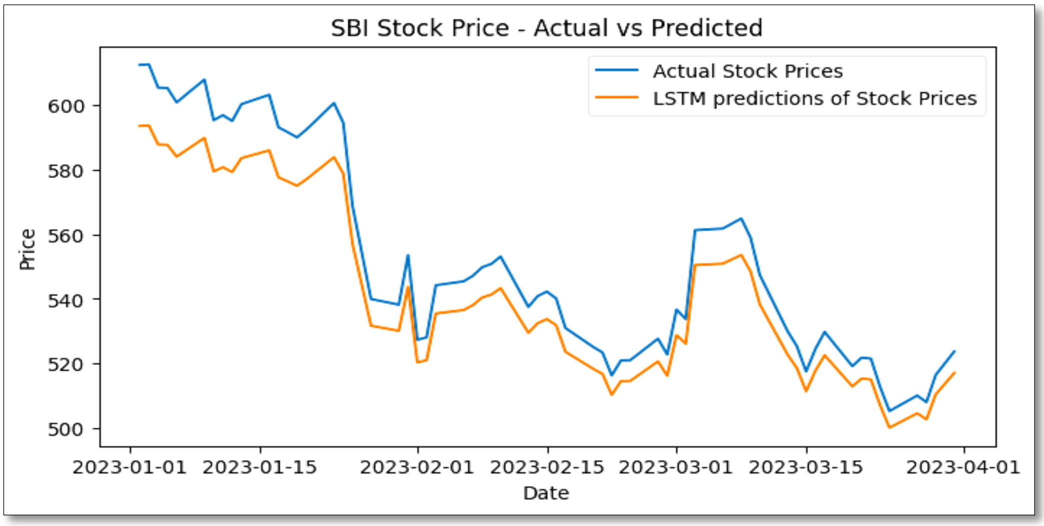

Figure 4(b): SBI BANK stock prices LSTM Predictions and Test graph

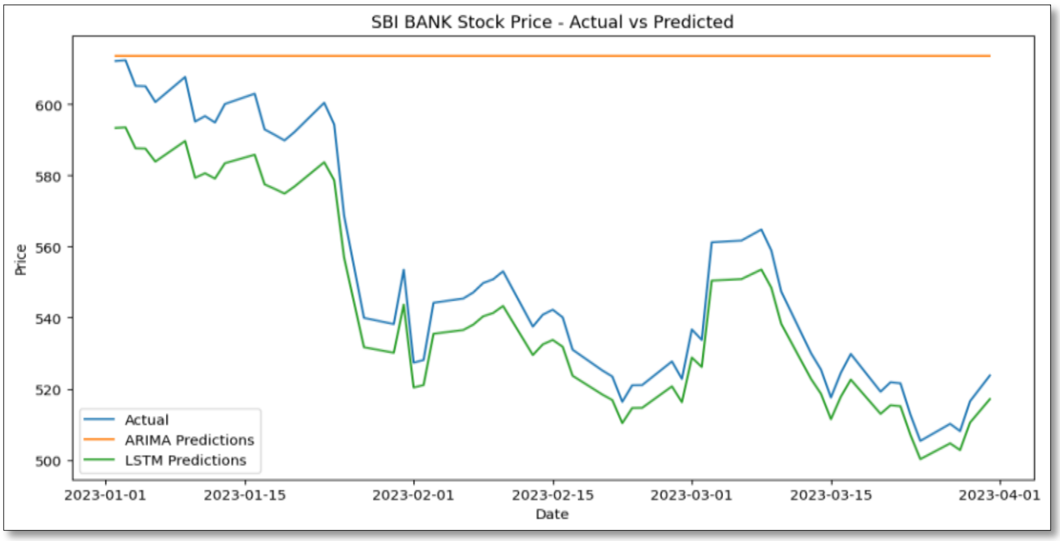

Figure 4(C): SBI BANK Stock Prices: LSTM Predictions ARIMA Predictions and Test data graph

TECHNIQUE	MAPE	RMSE	MAE	MSE
ARIMA(0,1,0)	11.5647	69.7892	61.7565	4870.5392
LSTM	1.8071	11.0310	10.1810	121.6820

Table 1: The Accuracy Matrix of the ARIMA Model and LSTM Model**.**

In Table 2, we have shown the actual test data set, predictions ARIMA model and LSTM model predictions in chronological order. These are also shown that LSTM predictions are very close to the actual test data set than ARIMA predictions.

Table 2: Predictions Table of ARIMA and LSTM Models

Date	Actual test observations	ARIMA Predictions	LSTM Predictions
02/01/2023	612.200012	613.700012	593.361633
03/01/2023	612.400024	613.700012	593.524231
04/01/2023	605.200012	613.700012	587.643616
05/01/2023	605.099976	613.700012	587.561523
06/01/2023	600.650024	613.700012	583.896362
-	-	-	-
-	-	-	-
-	-	-	-
24/03/2023	505.350006	613.700012	500.221008
27/03/2023	510.149994	613.700012	504.657043
28/03/2023	508.100006	613.700012	502.765137
29/03/2023	516.500000	613.700012	510.491730
31/03/2023	523.750000	613.700012	517.105530

V. CONCLUSION

ARIMA and LSTM models estimate SBI BANK closing stock prices from 01/01/2023 to 31/03/2023. MAPE and RMSE find the best ML model. LSTM's lower RMSE improves SBI BANK stock price prediction. Complex ML, particularly DL algorithms, has shown both models' flaws. LSTM and ARIMA models use time series relationships without considering additional parameters. Time, market, political, economical, industrial, and other factors and the company's management and structure affect the stock price. LSTM can employ indicators like RSI and MSCD to better predict market volatility and quick changes; however, it should be used with other models. LSTM and ARIMA models exploit time series relationships without considering external factors, highlighting their shortcomings. Time, market, political, economical, industrial, and other factors and the company's management and structure affect the stock price. Using indicators like RSI and MSCD, the LSTM model may predict market volatility and quick shifts.

In future, we will create a stronger hybrid model using standard and deep learning methods and sentimental analysis. We'll model low-frequency and high-dimensional data.

REFERENCES

[1] TEAM COUNTERPOINT. (2021, February 18). Apple iPhone Market Share: By Quarter. Counterpoint Research. https://www.counterpointresearch.com/apple-iphone-market-share-quarter/
[2] Apple Reports Third Quarter Results. (2022, July 28). Apple Newsroom. https: // www. apple. com/ newsroom/ 2022/07/ apple-reports-third-quarter-results/
[3] Barro, R. J. (1990). The Stock Market and Investment. Review of Financial Studies, 3(1), 115–131. https: // doi. org/ 10. 1093/rfs/3.1.115
[4] Lee, T. K., Cho, J. H., Kwon, D. S., & Sohn, S. Y. (2019). Global stock market investment strategies based on financial network indicators using machine learning techniques. Expert Systems with Applications, 117(228-242), 228–242. https://doi.org/10.1016/j.eswa.2018.09.005
[5] Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). A comparative analysis of forecasting financial time series using Arima, lstm, and bilstm. arXiv preprint arXiv:1911.09512.
[6] Qin, J., Tao, Z., Huang, S., & Gupta, G. (2021, March). Stock price forecast based on ARIMA model and BP neural network model. In 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) (pp. 426-430). IEEE.
[7] Hu, Y., Tao, Z., Xing, D., Pan, Z., Zhao, J., & Chen, X. (2020, August). Research on stock returns forecast of the four major banks based on ARMA and GARCH model. In Journal of Physics: Conference Series (Vol. 1616, No. 1, p. 012075). IOP Publishing.
[8] Cao, Z., Ding, X., & Tao, Z. (2022, September). News Detection for Recurrent Neural Network Approach. In 2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI) (pp. 794-797). IEEE.
[9] Stevenson, S. (2007). A comparison of the forecasting ability of ARIMA models. Journal of Property Investment & Finance.
[10] Zhu, T., Liao, Y., & Tao, Z. (2022). Predicting Google's Stock Price with LSTM Model. Proceedings of Business and Economic Studies, 5(5), 82-87
[11] Dev Shah, Haruna Isah, and Farhana H Zulkernine, "Stock Market Analysis: A Review and Taxonomy of Prediction Techniques", International Journal of Financial Studies, Vol. 7, No. 26, doi:10.3390/ijfs7020026 (2019).
[12] Min Wen, Ping Li, Lingfei Zhang, and Yan Chen, "Stock Market Trend Prediction Using High-Order Information of Time Series", IEEE Access, Vol. 7, pp.28299-28308 (2019).
[13] Jonathan Creighton, and Farhana H Zulkernine, "Towards Building a Hybrid Model for Predicting Stock Indexes", IEEE International Conference on Big Data (2017).
[14] Sahaj Singh Maini, and Govinda K, "Stock Market Prediction using Data Mining Techniques", Proc. Int.

Conf. Intelligent Sustainable Systems, IEEE Xplore Compliant - Part Number: CFP17M19-ART, pp.654-661 (2017).

[15] Mohammad Asiful Hossain, Rezaul Karim, Ruppa Thulasiram, Neil D B Bruce, and Yang Wang, “Hybrid Deep Learning Model for Stock Price Prediction”, IEEE Symposium Series on Computational Intelligence SSCI, pp.1837-1844 (2018).

[16] Dev Shah, Wesley Campbell, and Farhana H Zulkernine, “A Comparative Study of LSTM and DNN for Stock Market Forecasting”, IEEE International Conference on Big Data (2018).

[17] Mohit Iyer, and Ritika Mehra, “A Survey on Stock Market Prediction”, IEEE 5th International Conference on Parallel, Distributed and Grid Computing, Solan, India, pp.663-668 (2018).

[18] Jing Zhang, Shicheng Cui, Yan Xu, Qianmu Li, and Tao Li, “A Novel Data-driven Stock Price Trend Prediction System”, Elsevier Expert Systems with Applications, Vol. 97, pp.60–69 (2018).

[19] Sreelekshmy Selvin, Vinayakumar R, Gopalakrishnan EA, Vijay Krishna Menon, and Soman KP, “Stock Price Prediction using LSTM, RNN and CNN- Sliding Window Model”, IEEE Conference Paper, doi: 10.1109/ICACCI.2017.8126078, pp.1643- 1647 (2017).

[20] Luca Di Persio, and Oleksandr Honchar, “Recurrent Neural Networks Approach to the Financial Forecast of Google Assets”, International Journal of Mathematics and Computers in Simulation, Vol. 11, pp.7-13 (2017).

[21] K Hiba Sadia, Aditya Sharma, Adarrsh Paul, Sarmistha Padhi, and Saurav Sanyal, “Stock Market Prediction Using Machine Learning Algorithms”, International Journal of Engineering and Advanced Technology, Vol. 8, No. 4, pp.25-31 (2019).