Research Article

Stock price forecasting using Long Short Term Memory

[1,*]Tri Andi, [2]Candra Juni Cahyo Kusuma

1 Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia
2 Universitas PGRI Yogyakarta, Yogyakarta, Indonesia

* Corresponding Author: triandi@umy.ac.id

Abstract: The objective of this research is to develop a solution for predicting BRI stock prices using Long Short-Term Memory (LSTM) models. The LSTM model was selected for its capacity to process extensive time series data and discern latent temporal patterns. In this study, a BRI stock dataset obtained from Yahoo Finance is employed for the training and testing of an LSTM model. The evaluation results demonstrate that the LSTM model exhibits excellent predictive performance, with a mean absolute percentage error (MAPE) of 1.58768% and a root mean square error (RMSE) of 81.88216%. The Google test results demonstrate a low mean absolute percentage error (MAPE) of 1.5%, indicating a strong correlation between the predicted and true values. In other words, the RMSE values indicate the absolute error level in predictions, indicating the extent to which the model performs well when predicting a value that takes into account the context of the data. In conclusion, the proposed LSTM model shows promise for use in stock price prediction applications. The precision of these models can be tested by using them to make predictions, which would validate the decision-making supported by data. This research suggests that there is room for improvement of these models using techniques such as hyperparameter optimization or ensemble methods (bagging with other weak learners, etc.) to improve their accuracy.

Keywords: predicting, LSTM Model, optimization, accuracy, improvement

Citation: Tri Andi, & Candra Juni Cahyo Kusuma. (2025). Stock price forecasting using Long Short Term Memory. *Iota*, 5(1). https://doi.org/10.31763/iota.v5i1.900

Academic Editor: Adi, P.D.P

Received: January 01, 2025

Accepted: February 07, 2025

Published: March 08, 2025

1. Introduction

In the contemporary era, equities represent the most popular avenue for investment among a diverse array of investors, spanning the spectrum from individual to corporate entities. There is a plethora of publicly traded stocks in Indonesia, with Bank Rakyat Indonesia (BRI) representing one of the most liquid shares on the Indonesian Stock Exchange. As one of the largest banks in Indonesia, BRI is also regarded as a prime investment opportunity for both individual and institutional investors, as well as financial analysts. This will, of course, inform the decision-making process regarding the optimal investment strategy. Therefore, monitoring BRI share price movements is a crucial aspect of investment analysis [1].

Several factors influence the movement of a company's share price, including economic conditions, financial reports, and market sentiment. The prediction of stock prices is a complex problem due to the numerous factors that influence them. For many investors, predicting the dynamics of this rapidly evolving and unpredictable marketplace is a challenging endeavor. The impact of these dynamics on individual inventory prices is often unpredictable. It is therefore imperative to develop a method that can efficiently ingest historical data to predict stock price movements with greater consistency[2].

Another approach is the utilization of Long Short-Term Memory (LSTM), which may be familiar to those who prefer neural networks over other methodologies [3]. LSTM is capable of capturing the temporal dynamics of historical data, thereby facilitating more precise forecasting than traditional methods. The application of LSTM in financial data analysis has significant potential for stock price forecasting in finance [4].

Moreover, to implement LSTM in BRI share price predictions, it is necessary to gain a detailed understanding of the structure and dynamics of historical data. The data in the LSTM model should contain multiple related variables, such as close price, trade volume, and other technical indicators. The use of LSTM also enables more effective advice to be given and can lead to a reduction in startup risks in scaling the business[5]

It can be concluded that the use of an LSTM model for price prediction in the stock market provides an efficient and advanced solution for dealing with the inherent complexities of such markets [2], [6]. The LSTM is capable of considering a substantial amount of historical data, which makes it a highly valuable tool for investors seeking to enhance their knowledge and anticipate the movement of BRI share prices. This enhances the precision of the predictions while simultaneously assisting investors in avoiding the potential pitfalls associated with risky investment strategies.

2. Method

Subsequently, the LSTM model can be constructed following its design as a recurrent neural network architecture, thereby enabling the capture of the temporal progression of historical data.

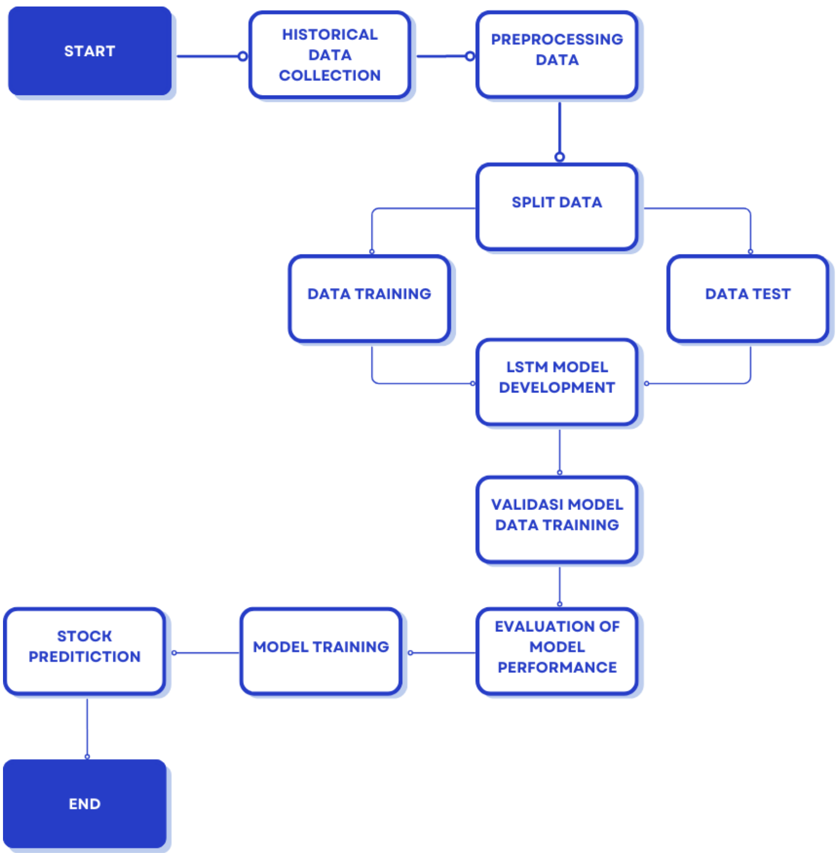

Figure 1. Research Methodology

The refined training data is then fed through the model to undergo training, with the training process itself involving the adjustment of hyperparameters to achieve optimal results. Once the model has been trained, it is validated against a separate set of test data, which indicates the accuracy of the resulting predictions. An illustration of this research method can be seen in the figure 1.

Two types of performance evaluation are conducted. The first is to ascertain the accuracy of predictions for range price data. This is done by comparing the predictions with actual stock prices. The second type of evaluation measures the number of incorrect predictions made. This is done using metrics such as Mean Average Percentage Error

(MAPE) and Root Mean Square Error (RMSE). Moreover, the model is evaluated in comparison to alternative prediction methods to ascertain which method demonstrates the superior predictive capability of stock prices relative to LSTM [7].

Subsequently, a more detailed analysis is conducted to identify the factors that contribute to the accuracy of predictions and to examine patterns in historical data, thereby enhancing the comprehension of the movements exhibited by stock prices. The findings of this study guide investors and financial analysts on the implementation of the LSTM model in stock market investment[8].

2.1 Data Collection

This research data uses BRI daily stock history data because BRI Bank is one of the largest banks in Indonesia, this data is taken from June 10, 2004, to June 10, 2023, sourced from finance.yahoo.com. The structure of BRI stock data available on Yahoo Finance can be seen in Table 1.

Table 1. BRI Stock Data Structure

Date	Open	High	Low	Close	Adj Close	Volume
10/06/2004	143.1794	145.452	143.1794	145.452	71.76071	178505583
11/06/2004	145.452	147.7247	143.1794	147.7247	72.882	53862430
14/06/2004	145.452	147.7247	143.1794	147.7247	72.882	95608151

The visual graph illustrates the fluctuating monetary value of stock belonging to Bank Rakyat Indonesia from June 2004 up until June 2023.

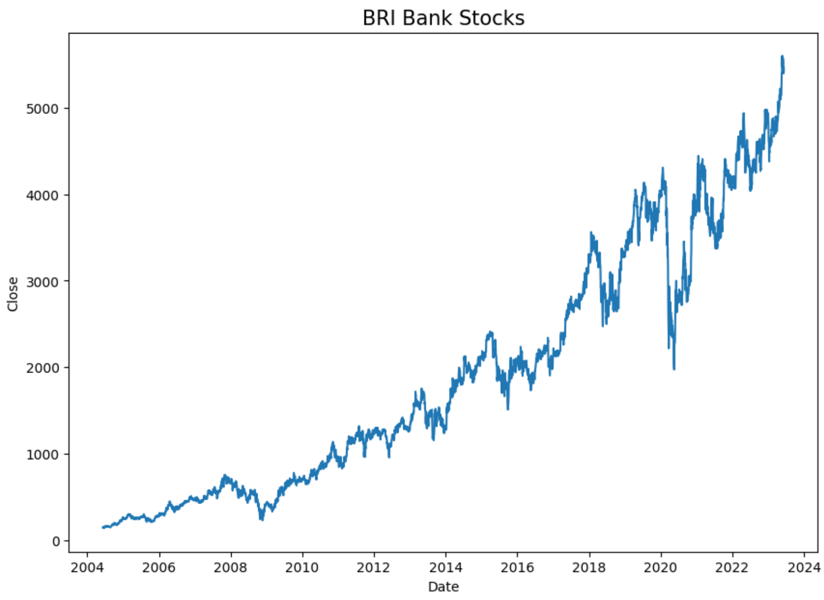

Figure 2. BRI Stock History chart

The illustration demonstrates an upward trajectory in the value of BRI shares across nearly two decades. From 2004 onwards, the company's stock price exhibited consistent growth with occasional fluctuations reflecting the dynamics of the market, as illustrated in Figure 2.

2.2 Data Processing

In this process, the downloaded data will be subjected to cleaning. This entails addressing issues about missing data, noisy data, and data transformation. Subsequently, the data will be partitioned into 80% training data and 20% test data. The data employed in this study is close data from BRI shares.

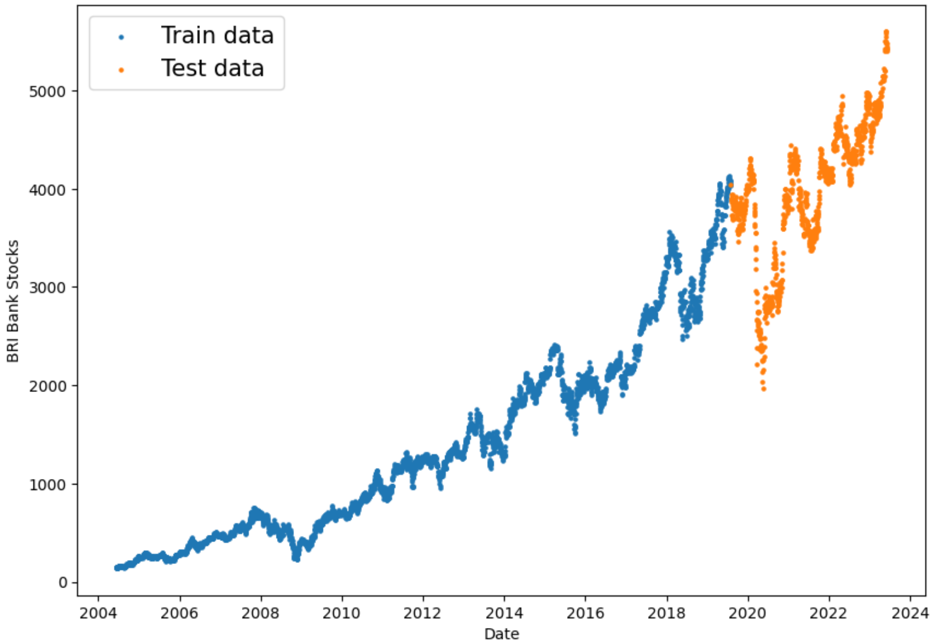

Figure 3. Division of Training Data and Testing

2.3 Model Training and Evaluation

This machine learning model uses LSTM. For this training process 80% of the data, and an epoch of 500. The Root Mean Squared Error (RMSE) and the Mean Average Percentage Error (MAPE) are employed to evaluate the precision of the model's prediction. The MAPE and RMSE can be calculated using the provided equations 1 & 2. [9], [10].

$$MAPE = \frac{1}{T}\sum_{i=1}^{T}\left|\frac{d_i - \hat{d_i}}{d_i}\right| \times 100 \quad [1]$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}\left(d_i - \hat{d_i}\right)^2} \quad [2]$$

3. Result and Discussion

The objective of the testing phase is to ascertain the efficacy of the identification process that has been constructed. The implementation of testing utilizes RMSE and MAPE, which facilitate a more comprehensive evaluation of the model's performance. The following section presents the RMSE and MAPE outcomes for the LSTM model that has been developed:

Table 2. RMSE and MAPE results of the LSTM model

Type	Results
RMSE	81.88216
MAPE	1.58768

It should be noted that the Mean Absolute Percentage Error (MAPE) is a metric that calculates the average error in percentage terms [11]. The MAPE value is calculated by taking the mean of the ratio between the absolute error and the actual value, expressed as a percentage[12], [13]. The mean absolute percentage error (MAPE) for the long short-term memory (LSTM) model is 1.58768, indicating that the average discrepancy between the predicted and actual values is approximately 1.58%.

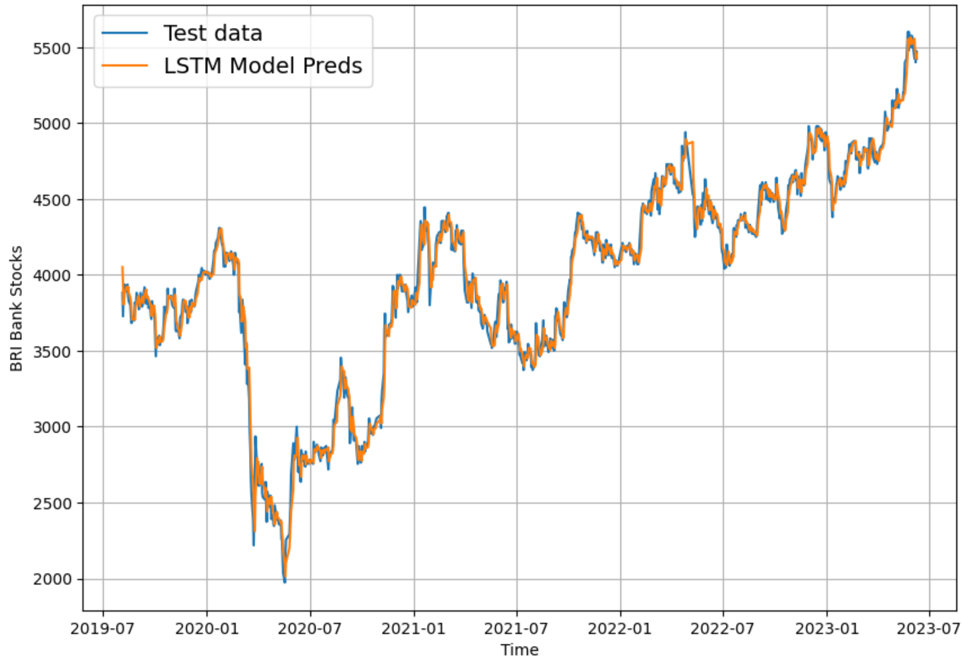

Figure 4. Comparison between test data graph and model prediction graph

This demonstrates that the LSTM model's prediction is highly accurate and that the error is also minimal in comparison to the original value. In contrast, the RMSE is a metric that quantifies the magnitude of the average discrepancy between the predicted and actual values [14]. The root mean square error (RMSE) is calculated as the square root of the mean of the squares of all differences between the predicted and actual values[15]–[20]. The RMSE for the long short-term memory (LSTM) model is 81.88216, which indicates that the average error between the predicted value and the actual target value is approximately 81.88 units. This provides insight into the magnitude of the error on the original scale of the data in comparison to the LSTM prediction model.

4. Conclusion

This study develops and evaluates the Long Short-Term Memory (LSTM) model to predict BRI stock prices. The evaluation results show that the model exhibits an excellent level of performance, as evidenced by the following error metrics: The average absolute percentage error (MAPE) is 1.58768%, while the root mean squared error (RMSE) is 81.88216. The low MAPE figure indicates that this LSTM model development is an accurate prediction, with an average error of about 1.58% of the actual value. This indicates that the modeling is reliable in the context of stock prices. The Root Mean Squared Error (RMSE) was obtained as 81.88216, which describes the absolute error rate. The RMSE value depends on the original scale of the data and the variability used in this study. Although the figure provides information about the share of error in the original data, its interpretation is highly dependent on the original scale and variability of the data used in this study. When compared to industry standards or results produced by alternative models, the RMSE provides a more detailed representation of the model's efficacy. Overall, this study shows that the LSTM model has significant potential as a tool for stock price prediction. The aim of applying this model is to improve the accuracy of predictions and provide practical benefits for decision-makers based on factual evidence. Future research can be done with the R2 test and cross-validation and will be compared with other methods optimized with PSO or CPSO using real-time data.

Acknowledgments: Thanks to all colleagues at Universitas Muhammadiyah Yogyakarta and Universitas PGRI Yogyakarta who have collaborated in completing this research, hopefully in the future, this research collaboration can continue to be improved and developed.

Author contributions: The authors are responsible for building Conceptualization, Methodology, analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision of project administration, funding acquisition, and have read and agreed to the published version of the manuscript.

Funding: The study was conducted without any financial support from external sources.

Availability of data and Materials: All data are available from the authors.

Conflicts of Interest: The authors declare no conflict of interest.
Additional Information: No Additional Information from the authors.

References

[1] M. Merfin and R. S. Oetama, "Prediksi Harga Saham Perusahaan Perbankan Menggunakan Regresi Linear Studi Kasus Bank BCA Tahun 2015-2017," Ultim. J. Tek. Inform., vol. 11, no. 1, pp. 11–15, Aug. 2019, doi: 10.31937/ti.v11i1.1239.
[2] L. Qin, K. Shanks, G. A. Phillips, and D. Bernard, "The Impact of Lengths of Time Series on the Accuracy of the ARIMA Forecasting," Int. Res. High. Educ., vol. 4, no. 3, p. 58, 2019, doi: 10.5430/irhe.v4n3p58.
[3] M. M. Frank et al., "Impact of Phase-Change Memory Drift on Energy Efficiency and Accuracy of Analog Compute-in-Memory Deep Learning Inference (Invited)," in IEEE International Reliability Physics Symposium Proceedings, 2023, vol. 2023-March. doi: 10.1109/IRPS48203.2023.10117874.
[4] D. Kumar Sharma, R. Prakash Varshney, S. Agarwal, A. Ali Alhussan, and H. A. Abdallah, "Developing a multivariate time series forecasting framework based on stacked autoencoders and multi-phase feature," Heliyon, vol. 10, no. 7, p. e27860, 2024, doi: 10.1016/j.heliyon.2024.e27860.
[5] Z. Alameer, M. A. Elaziz, A. A. Ewees, H. Ye, and Z. Jianhua, "Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm," Resour. Policy, vol. 61, no. September 2018, pp. 250–260, 2019, doi: 10.1016/j.resourpol.2019.02.014.
[6] Z. Chen, B. Wang, and A. N. Gorban, "Multivariate Gaussian and Student-t process regression for multi-output prediction," Neural Comput. Appl., vol. 32, no. 8, pp. 3005 – 3028, 2020, doi: 10.1007/s00521-019-04687-8.
[7] M. S. Khan and U. Khan, "Comparison of Forecasting Performance with VAR vs. ARIMA Models Using Economic Variables of Bangladesh," Asian J. Probab. Stat., no. December, pp. 33–47, 2020, doi: 10.9734/ajpas/2020/v10i230243.
[8] M. A. A. Bakar, N. Mohd Ariff, M. S. Mohd Nadzir, O. L. Wen, and F. N. A. Suris, "Prediction of Multivariate Air Quality Time Series Data using Long Short-Term Memory Network," Malaysian J. Fundam. Appl. Sci., vol. 18, no. 1, pp. 52 – 59, 2022, doi: 10.11113/MJFAS.V18N1.2393.
[9] D. C. Aishwarya and C. N. Babu, "Prediction of time series data using GA-BPNN based hybrid ANN model," in Proceedings - 7th IEEE International Advanced Computing Conference, IACC 2017, 2017, pp. 848 – 853. doi: 10.1109/IACC.2017.0174.
[10] L. Lian and Z. Tian, "A novel multivariate time series combination prediction model," Commun. Stat. - Theory Methods, vol. 53, no. 7, pp. 2253 – 2284, 2024, doi: 10.1080/03610926.2022.2124522.

[11] A. Famili, W.-M. Shen, R. Weber, and E. Simoudis, "Data preprocessing and intelligent data analysis," Intell. data Anal., vol. 1, no. 1, pp. 3–23, 1997.
[12] K. B. Debnath and M. Mourshed, "Forecasting methods in energy planning models," Renew. Sustain. Energy Rev., vol. 88, no. August 2016, pp. 297–325, 2018, doi: 10.1016/j.rser.2018.02.002.
[13] B. Prabawa, J. Nasri, and M. D. Sulistiyo, "Prediksi Harga Saham dengan Menggunakan Metode Autoregressive dan Algoritma Kelelawar," Jmm Unram - Master Manag. J., vol. 9, no. 2, pp. 107–121, 2020, doi: 10.29303/jmm.v9i2.508.
[14] A. Bulatov, "Sheridan College SOURCE : Sheridan Institutional Repository Forecasting Bitcoin Prices Using N-BEATS Deep Learning Architecture Forecasting Bitcoin Prices Using N-BEATS Deep Learning Architecture Sheridan College, Institute of Technology and Advanced Lea," 2020.
[15] R. H. Kusumodestoni and S. Sarwido, "Komparasi Model Support Vector Machines (Svm) Dan Neural Network Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham," J. Inform. Upgris, vol. 3, no. 1, 2017, doi: 10.26877/jiu.v3i1.1536.
[16] X. Wang, C. Li, C. Yi, X. Xu, and J. Wang, "Engineering Applications of Artificial Intelligence EcoForecast : An interpretable data-driven approach for short-term macroeconomic forecasting using N-BEATS neural network," Eng. Appl. Artif. Intell., vol. 114, no. June, p. 105072, 2022, doi: 10.1016/j.engappai.2022.105072.
[17] T. Khatib, A. Mohamed, and K. Sopian, "A review of solar energy modeling techniques," Renew. Sustain. Energy Rev., vol. 16, no. 5, pp. 2864–2869, Jun. 2012, doi: 10.1016/j.rser.2012.01.064.
[18] D. P. Anggriningrum, P. Hendikawati, and Z. Abidin, "Perbandingan Prediksi Harga Saham Dengan Menggunakan Jaringan Syaraf Tiruan Backpropagation Dan Arima," Unnes J. Math., vol. 2, no. 2, 2013, doi: 10.15294/ujm.v2i2.3249.
[19] B. P. Yafitra, Indwiarti, and A. R. Atiqi, "Perbandingan Prediksi Harga Saham Dengan Model Arima Dan Artificial Neural Network," Ind. J. Comput., vol. 4, no. 2, pp. 189–198, 2019, doi: 10.21108/indojc.2019.4.2.344.
[20] A. Arfan and L. ETP, "Perbandingan Algoritma Long Short-Term Memory dengan SVR Pada Prediksi Harga Saham di Indonesia," Petir, vol. 13, no. 1, pp. 33–43, 2020, doi: 10.33322/petir.v13i1.858.