IMPLEMENTATION OF LONG SHORT TERM MEMORY (LSTM) ALGORITHM FOR PREDICTING STOCK PRICE MOVEMENTS OF LQ45 INDEX (CASE STUDY: BBCA STOCK PRICE)

Cindi Rahayu, Dahlan Abdullah, Zara Yunizar
Informatics Engineering Universitas Malikussaleh[1,2,3]
Corresponding Email: cindi.rahayu@gmail.com

Abstract

.
This research aims to implement the Long Short Term Memory (LSTM) algorithm in predicting the movement of LQ45 stock prices. In this study, historical data of BBCA stock prices were used as an example of LSTM method implementation. The development process of the stock price prediction application begins with the collection of historical data, which then undergoes a preprocessing stage for normalization. The data is divided into training and testing sets, and transformed into suitable sequences for LSTM model input. The LSTM model is trained using the backpropagation through time algorithm and tested using the testing data. The predicted results from the LSTM model are compared with the actual labels using RMSE and MAPE metrics. Once satisfactory predictions are obtained, they are stored in a database and presented to users in the form of graphs and comparison tables. The implementation of LSTM in this research demonstrates prediction accuracy with an error percentage below 6%, with MAPE of 5.4772% and RMSE of 6.658%. Furthermore, the implementation of LSTM in the developed application using the latest historical data also yields low error percentages, with MAPE ranging from 3.7763% to 5.8048% for various stock price features. In conclusion, the LSTM method can be used for predicting stock price movements with satisfactory accuracy, providing valuable information for investment decision-making.

Keywords : LSTM, Stock, Prediction, Historical Data.

INTRODUCTION

Technological advancements have made financial transactions easier for society. From payments, ordering, to investments, everything can be done effortlessly through various smartphone apps, websites, and computers. Investments have become popular among different layers of society, including both upper and lower economic classes, as well as people of all ages. Investment is considered as a guarantee for achieving financial freedom in the future, prompting many young people to invest from an early age. (Jaen et al., 2019)However, there are also those who engage in rapid stock trading, such as buying and selling stocks in a short period. To predict future stock price movements, two analysis methods are used: fundamental analysis, which examines company fundamentals, and technical analysis, which employs statistics and mathematical approaches.(Khumaidi et al., 2020) This research will test the LSTM (Long Short Term Memory) method for predicting stock price movements. LSTM was chosen for its assistance in statistical calculations to predict historical data movements. The study aims to implement the LSTM algorithm in predicting the LQ45 stock price movements from 2017 to 2022. It is expected that this testing will provide insights into the accuracy of the method in predicting stock prices in the future.

LITERATURE REVIEW

Stock Market

Stocks in Indonesia Language are evidence of ownership in a company, and the term "saham" itself is derived from Arabic. In Islamic jurisprudence literature, the term "saham" is taken from the term "musahamah," which comes from the plural form "ashum" or "suhmah," meaning ownership share, indicating that shareholders are the owners of the company. The larger the number of shares owned, the greater the influence one has in the company. By issuing stocks, companies can raise long-term funds by selling ownership interests in the form of equity securities in exchange for cash. This is a primary method of raising business capital, alongside issuing bonds. Stocks are sold in the primary market or secondary market. Stocks are one of the most popular financial market instruments. Companies choose to issue stocks when seeking funding. On the other hand, stocks are a favored investment instrument for investors due to their potential for attractive returns.(Kurniasi et al., 2021)In Other Definition, Stocks are valuable documents that represent ownership shares in a company.

This means that when someone decides to buy stocks, they are essentially purchasing a portion of ownership in the company they are buying from. Another definition of stocks can be understood as a unit of value or bookkeeping in the financial component that focuses on the form of ownership in a company. So, when someone buys stocks in a company, they actually acquire rights to an asset and also earnings derived from the company according to the proportion of stocks they purchased. In simple terms, stocks are evidence of ownership in a company. The typical form of stocks is a piece of paper stating that the holder is the owner of the security issued by the company.(Syahputra, 2020)

LSTM (Long Shot Term Memory)

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture used for sequential data processing, such as time series, text, and speech. It is designed to overcome the limitations of traditional RNNs in capturing long-range dependencies and avoiding the vanishing/exploding gradient problem. (Ashari & Sadikin, 2020)LSTM is particularly effective in tasks where context and temporal patterns play a crucial role. The main idea behind LSTM is the inclusion of special memory units called "cells" that are equipped with three types of gates: the input gate, the forget gate, and the output gate. These gates regulate the flow of information within the LSTM cell, allowing it to selectively remember, forget, and output relevant information over time.(Simatupang et al., 2022)

The process of LSTM can be summarized in the following steps:

1. Input Step:
   - Input Gate: Decides which information from the current input should be stored in the cell state.
   - Forget Gate: Determines which information from the previous cell state should be forgotten.
   - Candidate Values: Calculate the candidate values that could be added to the cell state.
2. Update Cell State:
   - The cell state is updated by combining the results from the input gate, forget gate, and candidate values.
3. Output Step:
   - Output Gate: Decides what information from the updated cell state should be output as the final prediction or hidden state.

By using these gates, LSTM can effectively retain relevant information over long sequences and mitigate the problems of vanishing and exploding gradients. This makes it a powerful tool for various applications, such as time series forecasting, natural language processing, and speech recognition. (Rizal & Soraya, 2018)

Input Gate:

$i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + b_i)$

where:

i_t = *input gate activation at time step* t.

W_{xi} *and* W_{hi} = *weight matrices for the input and previous hidden state, respectively.*

x_t	= *input at time step* t.
h_{t-1}	= *previous hidden state.*
b_i	= *bias for the input gate.*
σ	= *sigmoid activation function.*

Forget Gate:

$$f_t = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + b_f)$$

where:

f_t	= *forget gate activation at time step* t.
W_{xf} and W_{hf}	= *weight matrices for the input and previous hidden state, respectively.*
x_t	= *input at time step* t.
h_{t-1}	= *previous hidden state.*
b_f	= *bias for the forget gate.*
σ	= *sigmoid activation function.*

Candidate Values:

$$\tilde{C}_t = tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_c)$$

where:

$\tilde{C}_t$	= *candidate value at time step* t.
W_{xc} and W_{hc}	= *weight matrices for the input and previous hidden state, respectively.*
x_t	= *input at time step* t.
h_{t-1}	= *previous hidden state.*
b_c	= *bias for the candidate values.*
tanh	= *hyperbolic tangent activation function.*

Update Cell State:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

where:

C_t	= *cell state at time step* t.
f_t	= *forget gate activation at time step* t.
C_{t-1}	= *previous cell state.*
i_t	= *input gate activation at time step* t.
$\tilde{C}_t$	= *candidate value at time step* t.
$\odot$	= *denotes element-wise multiplication.*

Output Gate:

$$o_t = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + b_o)$$

where:

o_t	= *output gate activation at time step* t.
W_{xo} and W_{ho}	= *weight matrices for the input and previous hidden state, respectively.*
x_t	= *input at time step* t.
h_{t-1}	= *previous hidden state.*
b_o	= *bias for the output gate.*
σ	= *sigmoid activation function.*

METHOD

This research is conducted in the city of Lhokseumawe, precisely at Malikussaleh University, scheduled to start in December 2022 until completion. The study aims to test the accuracy of stock price movements using the LSTM method and also incorporates real-time stock price updates using the Stock Market API (api.marketstack.com). In this section, we will outline the sequential stages involved in the developed system.

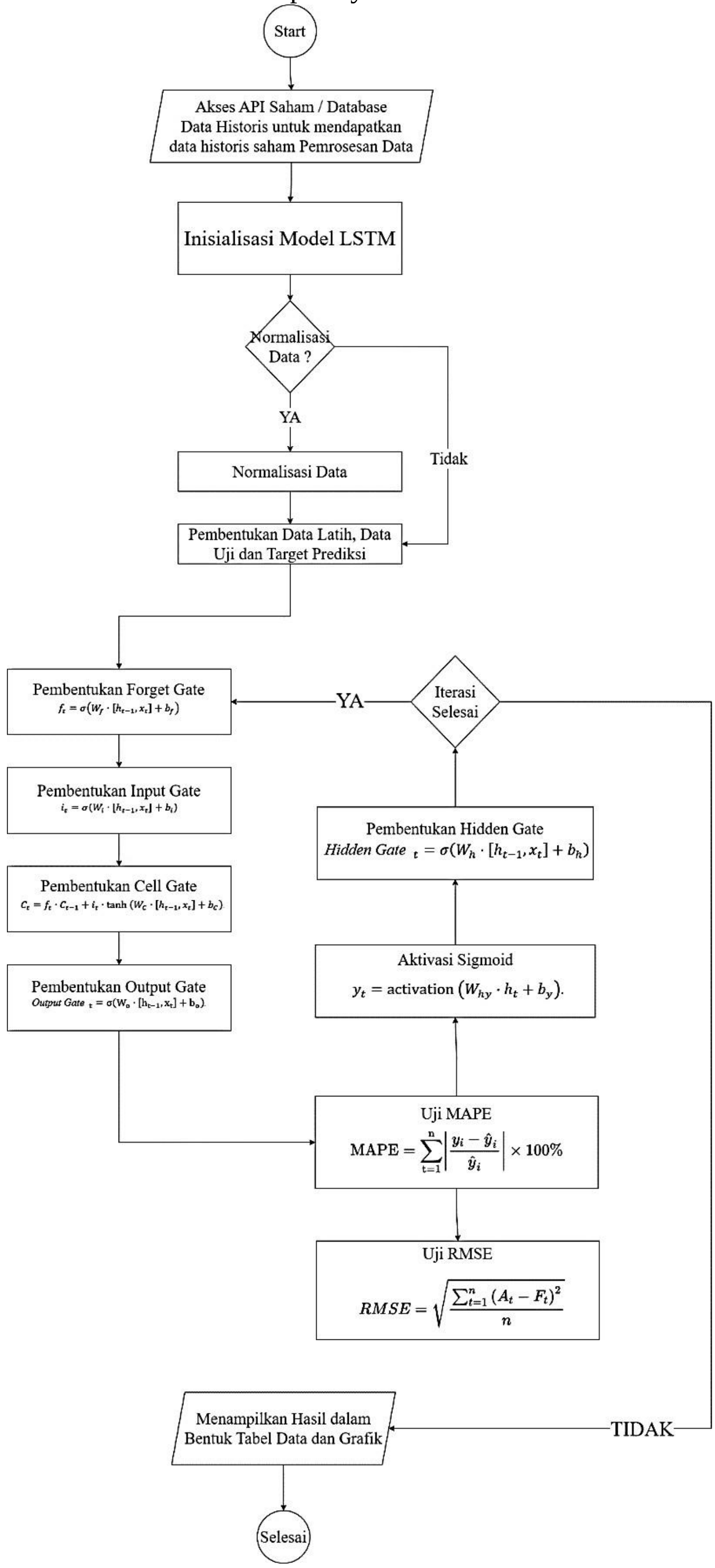

Image 1 System Flowchart

1. The application will access the API for Current Stock Prices, especially for LQ45, to obtain the latest data.
2. The application will retrieve historical data to access Historical Stock Data.
3. The application will initiate the Prediction process using LSTM:
    a. Initialization of the Training Model where several artificial intelligence training functions will be executed.
    b. The system will normalize the data obtained to reduce computational burden.
    c. Formation of Gates to calculate and predict stock price movements.
    d. Comparison of MAPE (Mean Absolute Percentage Error) calculations between the actual prices and the predicted prices.
    e. Checking the results between predicted prices, actual prices, and targets. If they do not match, the model will undergo iterative processing.
4. The application will store the prediction results in the database.
5. The application will display the prediction results in the form of tables and graphs.

The following data presents the historical stock prices used in this study, sourced from PT Bank BCA (BBCA) and spanning from January 1, 2020, to January 14, 2020. The dataset includes essential information such as Date, Open, High, Low, and Close prices. These records have been collected for research purposes and will be utilized in the analysis and evaluation of the stock's performance during the specified period.

Table 1 BBCA Historical Data

Date	High	Low	Open	Close
1/1/2020	Rp31,550.00	Rp31,000.00	Rp31,200.00	Rp31,500.00
2/1/2020	Rp32,000.00	Rp31,400.00	Rp31,600.00	Rp31,550.00
3/1/2020	Rp31,500.00	Rp31,100.00	Rp31,400.00	Rp31,300.00
6/1/2020	Rp32,000.00	Rp31,500.00	Rp31,550.00	Rp31,700.00
7/1/2020	Rp32,200.00	Rp31,600.00	Rp31,700.00	Rp31,800.00
8/1/2020	Rp32,000.00	Rp31,400.00	Rp31,800.00	Rp31,600.00
9/1/2020	Rp31,800.00	Rp31,500.00	Rp31,600.00	Rp31,750.00
10/1/2020	Rp32,000.00	Rp31,600.00	Rp31,750.00	Rp31,900.00
13/1/2020	Rp32,100.00	Rp31,800.00	Rp31,900.00	Rp32,000.00
14/1/2020	Rp31,950.00	Rp31,600.00	Rp32,000.00	Rp31,850.00

The provided data has been normalized using the normalization formula. Normalization is a process that scales the data to a uniform range, typically between 0 and 1, to facilitate accurate comparisons and calculations during the analysis. By normalizing the data, any potential discrepancies in scale are eliminated, ensuring that each data point contributes equally to the overall analysis. This step ensures that the data remains consistent and suitable for further processing, allowing for meaningful insights and conclusions to be drawn from the research.

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

where:

X_{norm} = normalized value of the data point X.

X = original data point.

$X_{\min}$ = minimum value in the dataset.
$X_{\max}$ = maximum value in the dataset.

And Then Result Table Is :

Table 2 Normalize Data

Date	High	Low	Open	Close
1/1/2020	0.970	0.000	0.235	0.706
2/1/2020	1.000	0.250	0.471	0.735
3/1/2020	0.943	0.100	0.412	0.647
6/1/2020	1.000	0.400	0.500	0.824
7/1/2020	1.030	0.500	0.588	0.882
8/1/2020	1.000	0.250	0.647	0.765
9/1/2020	0.897	0.400	0.706	0.794
10/1/2020	1.000	0.500	0.765	0.824
13/1/2020	1.015	0.600	0.824	0.853
14/1/2020	0.985	0.500	0.853	0.824

RESULTS AND DISCUSSION

The following is a detailed presentation of the results derived from the LSTM computation process, encompassing the intricate steps of gating, activation, and the final outcome. In this section, we unveil the comprehensive outcomes obtained through the LSTM calculation process, which encompasses multiple crucial stages. Starting from the initial gate operations, where data is selectively processed and transformed, to the subsequent activation functions that regulate the flow of information and memory retention. Throughout this meticulous process, the LSTM algorithm effectively captures long-term dependencies and patterns within the data, leading to a powerful predictive capability. Finally, after undergoing these intricate steps, the LSTM model produces its ultimate output, providing valuable insights and predictions for further analysis and decision-making.

Table 3 Result After LSTM

Date	High	Low	Open	Close	Prediction
1/1/2020	0.6774	0.5547	0.5616	0.6639	0.5836
1/2/2020	0.6856	0.5719	0.5667	0.6675	0.5857
1/3/2020	0.6697	0.5491	0.5578	0.6607	0.5836
1/6/2020	0.7018	0.5904	0.5802	0.6797	0.5938
1/7/2020	0.709	0.5981	0.5865	0.6836	0.5969
1/8/2020	0.7071	0.597	0.5857	0.6825	0.5964
1/9/2020	0.6954	0.5832	0.5764	0.6757	0.5919
1/10/2020	0.7097	0.5988	0.5872	0.6839	0.5971
1/13/2020	0.7116	0.601	0.5892	0.6853	0.598

Below are the results of MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Squared Error) from the accuracy testing of the predictions. In this section, we present the outcomes of the accuracy evaluation conducted on the predictions made by the model. The MAPE metric provides insights into the average percentage difference between the predicted values and the actual values. A lower MAPE indicates a higher accuracy in the model's predictions. On the other hand, the RMSE metric measures the root mean squared difference between the predicted and actual values, giving an indication of the magnitude of prediction errors. The lower the RMSE value, the better the model's performance. These evaluation metrics serve as valuable tools to assess the predictive capabilities and precision of the developed system, aiding in the validation of its reliability and effectiveness.

Table 4 RMSE Prediction Result

Date	High	Low	Open	Close	Prediction	Diff	% Error
1/1/2020	0.970	0.000	0.235	0.706	0.734	0.027	2.79%
2/1/2020	1.000	0.250	0.471	0.735	0.740	0.004	0.52%
3/1/2020	0.943	0.100	0.412	0.647	0.726	0.079	12.26%
6/1/2020	1.000	0.400	0.500	0.824	0.759	0.065	7.87%
7/1/2020	1.030	0.500	0.588	0.882	0.765	0.118	13.33%
8/1/2020	1.000	0.250	0.647	0.765	0.763	0.002	0.29%
9/1/2020	0.897	0.400	0.706	0.794	0.750	0.044	5.58%

Table 5 MAPE Prediction Result

Date	Target (Close)	Prediction	Diff	% Diff
01/01/2020	0,735	0,734	-0,001	-0,1361%
02/01/2020	0,647	0,740	0,093	14,3740%
03/01/2020	0,824	0,726	-0,098	-11,8932%
06/01/2020	0,882	0,759	-0,123	-13,9456%
07/01/2020	0,765	0,765	0,000	0,0000%
08/01/2020	0,794	0,763	-0,031	-3,9043%
09/01/2020	0,824	0,750	-0,074	-8,9806%
10/01/2020	0,853	0,765	-0,088	-10,3165%
13/01/2020	0,897	0,767	-0,130	-14,4928%
MAPE				**5,4772%**

The implementation of LSTM in this research yields several conclusions. Among the 14 data points used, spanning from January 1, 2020, to January 14, 2020, with historical data from BBCA, the training data was taken from January 1, 2020, to January 13, 2020. The data extracted includes Open, Close, High, and Low values from the BBCA historical data. The process involves the use of weights and bias W_f=[0.2,0.3,0.4,0.1] and b_f=0.5 on the normalized data. After passing through the input gate, Dot Product, Cell Gate, Output Gate, Hidden Gate, and sigmoid activation, the RMSE result is 6.658%. The comparison between target predictions and actual predictions results in an MAPE of 5.4772%. The comparison between test data and predicted data shows a difference of -27.4242%, where the test data is 0.824 on January 14, 2020, while the prediction result is 0.598. Furthermore, in the implementation of LSTM in the developed application, using the latest historical data from May 25, 2023, to June 23, 2023, the MAPE for Close price is 4.8747%, Open price is 4.3686%, High price is 5.8048%, and Low price is 3.7763%. It can be concluded that LSTM predictions for the stock price of BBCA have an error percentage below 6%. For detailed results, please refer to the appendix, which presents tables for each company in the LQ45 list.

LSTM Complexity

Additionally, the following graph illustrates a comparison between the actual stock prices and the predicted values. In this section, we present a visual representation of the comparison between the real stock prices and the corresponding predicted values generated by the model. The graph allows for a direct and intuitive assessment of the model's performance in capturing the actual price trends. A close alignment between the actual and predicted lines signifies a high level of accuracy in the model's predictions, indicating its proficiency in capturing the underlying patterns and trends in the stock market data. This graphical representation aids in providing a comprehensive understanding of the model's predictive capabilities and contributes to the overall evaluation of its effectiveness in forecasting stock price movements.

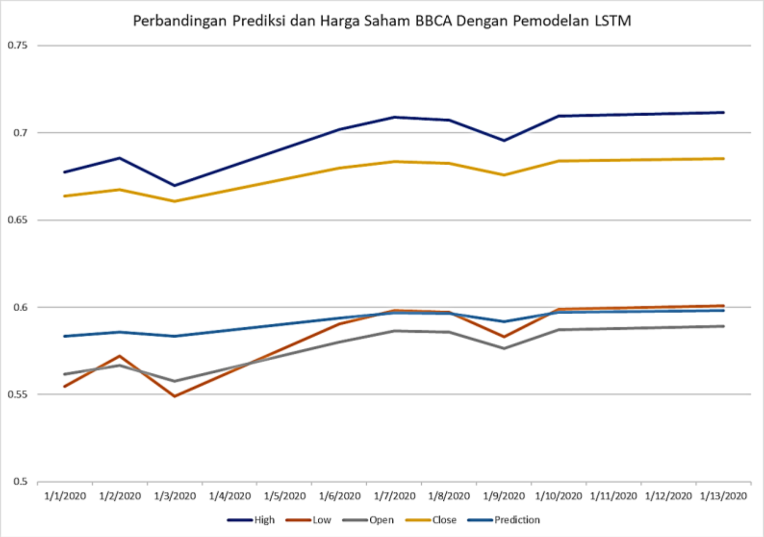

Image 2 Line Chart Real Price VS Prediction Result

This section outlines the essential steps involved in forming and training the LSTM model, including data collection, model formation, and parameter optimization. These computations demand significant computational power due to the complex structure and numerous parameters, which must be fine-tuned to achieve optimal performance. LSTM's complexity arises from its specialized architecture, involving input gates, forget gates, cell states, and output gates. With more parameters and intricate gradient calculations during training, LSTM requires substantial computation and precise parameter tuning. Despite this complexity, LSTM's ability to process sequential data and retain long-term information makes it invaluable in various applications.

Web Based LSTM Prediction Apps

Below is the application that has been developed using HTML, PHP, MYSQL, JS, and Tensorflow.js. It results in a web-based platform that displays stock price predictions. The application utilizes a combination of these programming languages and frameworks to create a dynamic and interactive web interface. It collects historical stock data from a MYSQL database, which is then processed using Tensorflow.js, a JavaScript library for machine learning. The predictive model is trained using the historical data, enabling it to forecast future stock prices. The web interface provides users with an intuitive platform to input stock symbols or select predefined stocks. The application then processes the input, runs the LSTM-based predictive model, and displays the forecasted stock prices in the form of charts and tables. This enables users to visualize the predicted trends and make informed decisions based on the insights provided by the application.

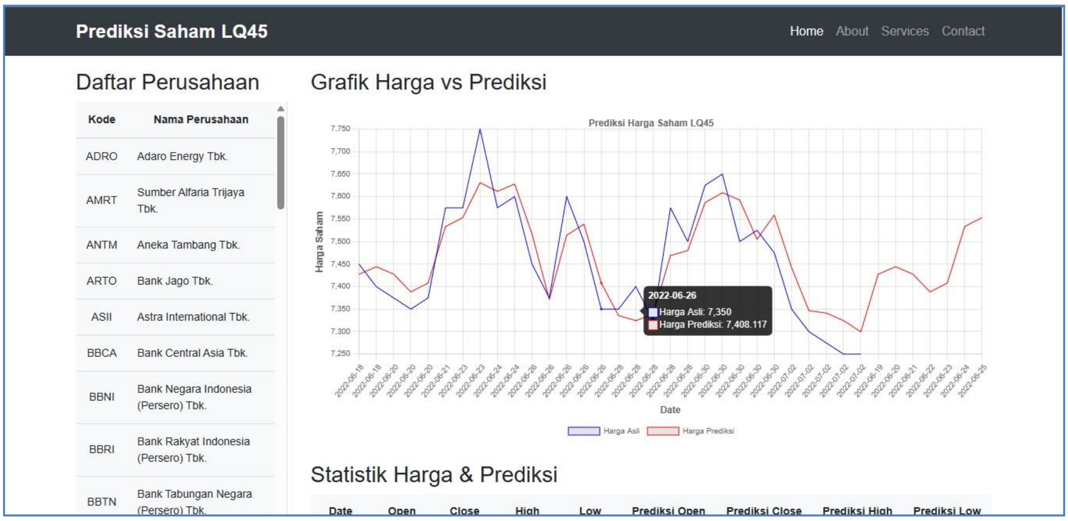

Image 3 Website For LSTM Prediction

CONCLUSION

Based on the conducted research and testing regarding the, the following conclusions can be drawn:

a. This application is carefully designed using visual modeling languages such as DFD, ERD, and context diagrams. The design process involves a step-by-step analysis using sample data from BBCA to implement LSTM methods in stock price prediction. Historical data from BBCA is utilized to build the prediction model, with important features and LSTM methods selected through visual modeling. Historical data from various stocks worldwide is obtained via the stackmarket.com API. After a comprehensive analysis, the application is developed using LSTM and BBCA sample data. The development process includes data collection, preprocessing, normalization, data division into training and testing sets, and preparing sequences suitable for the LSTM model. The LSTM model, composed of interconnected layers, is trained using the backpropagation through time algorithm to optimize predictions. Once trained, the LSTM model is tested using the testing data, and the results are evaluated using RMSE and MAPE metrics. Satisfactory predictions are stored in the database and presented to users through graphs and tables for valuable investment decision-making information.

b. The results of the LSTM implementation in this research lead to several conclusions. Utilizing 14 data points from January 1, 2020, to January 14, 2020, with historical data from BBCA, the training data is taken from January 1, 2020, to January 13, 2020, including Open, Close, High, and Low data. During the process, weight and bias values of W_f=[0.2,0.3,0.4,0.1] and b_f=0.5 are used on normalized data. After going through the input gate, Dot Product, Cell Gate, Output Gate, Hidden Gate, and sigmoid activation, an RMSE result of 6.658% is obtained. The comparison between target predictions and actual predictions yields an MAPE of 5.4772%. Additionally, the comparison between test data and predicted data shows a difference of -27.4242%, where the test data is 0.824 on January 14, 2020, while the prediction result is 0.598. Furthermore, in the implementation on the developed application, using the latest historical data from May 25, 2023, to June 23, 2023, the MAPE for Close price is 4.8747%, Open price is 4.3686%, High price is 5.8048%, and Low price is 3.7763%. The LSTM predictions exhibit an error percentage below 6% for the BBCA stock price prediction. The complete results can be found in the appendix, which includes tables for each company in the LQ45 list.

REFERENCES

Ashari, M. L., & Sadikin, M. (2020). Prediksi Data Transaksi Penjualan Time Series Menggunakan Regresi Lstm. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, *9*(1), 1. https://doi.org/10.23887/janapati.v9i1.19140

Jaen, H., Darnila, E., & Fikry, M. (2019). Aplikasi Peramalan Kurs Bitcoin-Rupiah Dengan Menggunakan Metode Double Exponential Smoothing. *TECHSI - Jurnal Teknik Informatika*, *11*(1), 106. https://doi.org/10.29103/techsi.v11i1.1295

Khumaidi, A., Raafi'udin, R., & Solihin, I. P. (2020). Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung. *Jurnal Telematika*, *15*(1), 13–18. https://journal.ithb.ac.id/telematika/article/view/340

Kurniasi, A. A., Saptari, M. A., & Ilhadi, V. (2021). Aplikasi Peramalan Harga Saham Perusahaan Lq45 Dengan Menggunakan Metode Arima. *Sisfo: Jurnal Ilmiah Sistem Informasi*, *5*(1), 13–26. https://doi.org/10.29103/sisfo.v5i1.4849

Rizal, A. A., & Soraya, S. (2018). Multi Time Steps Prediction dengan Recurrent Neural Network Long Short Term Memory. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *18*(1), 115–124. https://doi.org/10.30812/matrik.v18i1.344

Simatupang, C. G. K., Swastika, W., & Suganda, T. R. (2022). Perancangan Aplikasi Berbasis Web Untuk Prediksi Harga Saham Dengan Metode Lstm. *Sainsbertek Jurnal Ilmiah Sains & Teknologi*, *3*(1), 1–8. https://doi.org/10.33479/sb.v3i1.212

Syahputra, A. (2020). Implementasi Metode Double Exponential Smoothing Pada Aplikasi Prediksi Harga Emas Dosen Pendidikan Vokasional Teknik Mesin. *Jurnal Teknologi Terapan & Sains*, *1*, no.*1*, 12.