International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-84, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Prediction of Stock Price Volatility Using the Long Short Term Memory (LSTM) Model for Investment Portfolio Selection Strategy

Ana Khofifahturizqi[1], Farikhin[2], Ratna Herdiana[3], Yola Ulrike Sihombing[4]
[1,2,3,4] Department of Mathematics, Faculty of Sciences and Mathematics, Diponegoro University, Indonesia

ABSTRACT: Volatility is an important variable in financial data models. Predicting volatility in financial data is helpful for investors to make good decisions to reduce risk and to gain investment returns. In predicting volatility, many researchers have conducted research in building prediction models using data mining. This research uses a deep learning algorithm, namely Long Short Term Memory (LSTM) which has high accuracy compared to other models. The research aims to predict stock price volatility and for investment portfolio selection. The object of this study is the historical stock price of PT. Unilever Indonesia Tbk. (UNVR), PT. Fast Food Indonesia Tbk. (FAST) which manages KFC and PT. MAP Boga Adiperkasa Tbk. (MAPB) which manages Starbucks in the period 2023 to 2024, when there was a boycott caused by the war between countries that occurred in the Middle East. The data is analysed using the LSTM model where stock price volatility was determined by the variance of the return and log return on the next seven days, then using LSTM the stock price volatility data was predicted. The results show that the MSE and RMSE values are very small, which means that the volatility prediction results are almost the same as the actual data. And the average volatility prediction results in UNVR stock of 0.00841, MAPB stock of 0.01717, and FAST stock of 0.01323. From these results can be used as a reference for the selection of investment portfolios.

KEYWORDS: Investment Portfolio, Investment, LSTM, Stock Price, Volatility.

INTRODUCTION

It has been generally recognized by many researchers that financial volatility is referred to as a key factor to assessing the risk of financial assets. Volatility itself strongly influences consumers investment behavior and returns, potentially posing important risk to the stability of the financial system [1], [2]. Volatility is an important variable in financial data models, although volatility is not the same as risk, volatility is defined as the uncertainty of asset returns and is used as a key input by many investment and portfolio decisions [3]. Investors use the price volatility of the underlying asset as a tool to measure risk [4]. The higher the financial volatility, the wider the range of fluctuations in the value of financial assets and the greater the uncertainty around asset returns.

Volatility prediction in the field of financial data is very helpful for investors to make good decisions to reduce risk and gain investment benefits. Precise and accurate volatility prediction is important in effectively determining the derivative value of an item and hedging the underlying asset. Therefore, volatility in financial markets plays an important role for derivatives pricing, portfolio management and hedging strategies [5].

In predicting volatility, experts have done a lot of research in building prediction models using data mining. Many algorithms in deep learning are focused on teaching representational (non-linier) data at a high level. One of the effective deep learning algorithms for time series data prediction is the Long Short Term Memory (LSTM) algorithm which is a development of Recurrent Neural Network (RNN) [6]. According to Liu et al [7], the performance of LSTM is clearly superior to GARCH, in terms of MAE or MSE.

The research on the use of LSTM in prediction conducted by Beniwal et al [8] where six different prediction methods were compared including Deep Neural Network (DNN), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Bidirectional Short Term Memory (Bi-LSTM), Gated Recurrent Unit (GRU), and Convolutional Neural Network (CNN), The results of RMSE and MAPE his research show the superiority of LSTM for predicting long-term daily prices. Other than that, Petrozzielo et al [9] showed his research results on the advantages of LSTM over the well-known univariate R-GARCH and GJR-MEM methods, when forecasting and high volatility conditions are present. It can be concluded that LSTM can be used to predict daily stock price volatility.

This study aims to predict stock price volatility using LSTM and investment portfolio selection strategies from several stocks affected by the boycott due to conflicts that occurred in the Middle East in the period 2023 to 2024. From this incident, many stocks experienced a drastic decline including PT. Unilever Indonesia Tbk. (UNVR), PT. Fast Food Indonesia Tbk. (FAST) which manages KFC and PT. MAP Boga Adiperkasa Tbk. (MAPB) which manages Starbucks.

Many important events have a lot of influence on stock price volatility. Looking at previous studies that have conducted research on how political events, disasters and similar events can affect the market response on stock price volatility [10], [11]. The occurrence of these events tends to cause instability in stock price volatility in the market. Stock price volatility in the market, which can provide a great challenge to the prediction of stock price volatility [12]. Maulana [13] in his research explains the fact that in the Indonesian stock market volatility often increases rapidly when bad news appears compared to changes in volatility when there is good news in the sectoral indices. In the investment portfolio selection strategy when viewed from the value of volatility according to Chotib and Huda [14] in his research revealed that if volatility is high, the investment risk is higher. Short-term traders prefer volatility that has a high value because it is expected to be able to get a large return, usually in daily or weekly time. In contrast, an investor who wants stability of returns will favor minimal volatility, even if he has to hold his shares for a long period of time earn capital gains or profits.

BASIC CONCEPT

A. Volatility

According to Atkins et al [15] and García-Medina et al [16] in his research revealed the prediction of future volatility can be constructed using the average variance of the log return at each interval, and calculated by the ratio of closing prices at a point in time that are close to each other:

$$r_t = log\left(\frac{S_t}{S_{t-1}}\right) \tag{1.1}$$

$$\sigma^2 = \frac{1}{n}\sum_{t=1}^{n}(t_i - \bar{t}\,)^2 \tag{1.2}$$

where S_t is referred to as the stock price at time t, r_t is the asset return at time t and $\bar{t}$ the average value of the return over the analysis window of *n* time points.

B. Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) was first proposed by Hochreiter and Schimdhuber in 1997 [17]. According to Xu et al [12] LSTM is able to improve long-term memory using a logical control unit of three types of gates, including forgetting gates, inputs gates, and output gates that are used to update and store information in cell states, generally able to overcome the problem of missing and exploding gradients. The LSTM model is able to capture long-term dependencies in time series data, which leads to increased convergence and learning speed. This feature can help in improving the prediction accuracy of fluctuations in high frequency.

Three types of gates in the LSTM:

1. Forget gate

This gate is able to play an important role in determining what information should be kept or discarded from the cell state at time – 1. The forget gate receives input from a hidden layer state $h^{(t-1)}$ and the current input sequence x^t at time t, then outputs a value of value between 0 and 1. A value of 1 represents that all previous information is kept, and a value of 0 describe nothing is stored from the previous cell state. In this method, the LSTM effectively manages the flow of information and protects the loss of important data during time series analysis.

$$f^t = \sigma(x^t W_f + h^{(t-1)} U_f + b_f) \tag{1.3}$$

where:

f^t : represents the forgotten door at time *t*

W_f : weight parameter

U_f, b_f : bias term

σ : sigmoid function, that governs the flow of information.

Definitely, these parts work together in determining which information should be kept or forgotten/discarded as seen from the previous time step when the new.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-84, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

2. **Input gate**

In this type of gate, the input gate receives new information x^t and determines which one should be stored in the cell state. The memory cell state $\left(\tilde{C}(t)\right)$ is then calculated with a hyperbolic tangent layer. Its value is used as a step to update the cell state resulting in a new cell state C^t. The update process component of the input gate is described by equation:

$$i^t = \sigma(x^t W_i + h^{t-1} U_i + b_i) \tag{1.4}$$

$$\tilde{C}t = \tanh(W_c[h^{t-1}, x^t] + b_c \tag{1.5}$$

$$C^t = f^t \otimes C^{t-1} + i^t \otimes \tilde{C}t \tag{1.6}$$

where:

t : time
i^t : the input gate at time *t*, which regulates the flow of information from x^t to *C*
W_i : weight parameter
U_i, b_i : bias term
W_c and b_c : adjust the weight matrix and bias terms in the cell state
$tan(h)$: activation function the hyperbolic tangent
$\otimes$: hadamard product which is used for mathematical formulas.

3. **Output gate**

This gate is used to ascertain which information from the current cell state should be selected as output. The sigmoid layer controlling the output gate is able to decide which memory cell state should be exported, and the $tanh$ layer. It then processes the information to be selected as a result of the final output value o^t. Furthermore, the output value o^t is forwarded for input to the next hidden layer in the network. The equation below explains how the gate activation output will be calculated.

$$o^t = \sigma(x^t W_o + h^{t-1} U_o + b_o) \tag{1.7}$$

$$h^t = o^t \otimes tanh\, C^t \tag{1.8}$$

where:

h^t : the hidden state at this time
o^t : the output gate
t : time
W_o : the weight matrix
U_o, b_o : bias terms

C. Investment Portfolio

Markowitz first proposed the idea of the efficient frontier in 1952. Called the initial mathematical formula of portfolio optimization, his theory states that an investor sees variation in returns as undesirable. However, the investor views expected returns as favorable. Markowitz proved that there is an ideal set of portfolios for a given value of risk, and that it maximizes the expected return and I referred to as the ideal limit [18]. The investment strategy relates to suitable investment instruments, effective portfolio management, and controlling investment risk [19].

RESEARCH METHODOLOGY

This research was conducted using daily stock price data from several equities affected by the conflict between countries in the Middle East, the stocks included in the data are UNVR, MAPB, and FAST listed on the Indonesia Stock Exchange (IDX), and the data is obtained directly from the official yahoo finance website.

<u>The steps of data analysis are as follows:</u>

1. Data Preprocessing, the data used for the study is historical stock price data, namely Closing Price data of stock affected by the war between Middle East countries. The time series data used starts from the period 2023-05-11 to 2024-05-08. In the initial preprocessing stage, it is important to cleaning the data.
2. Determine the volatility to be predicted using return variance and log return, with the data used being the closing price.
3. Normalized stock price volatility data using MinMaxScaler.
4. Dividing the data into two parts with a ratio of 80% training data and 20% test data.

5. Build the LSTM model to predict volatility. With the parameters used taken from previous research results.

Table 1. Hyperparameter for LSTM model

Hyperparameter	**Value**
Input	Stock price volatility data and time
Data Partition	80% training data and 20% testing data
Dropout	0.6
Batch Size	25
Epoch	50
Hidden Layer	3
Optimizer	Adam

The above hyperparameters are used to allow the model to capture temporal dependencies and non-linier relationships in stock price volatility data, as well as reduce complex trends, and the resulting output is a prediction of stock price volatility.

6. Predicting stock price volatility.
7. Perform model evaluation using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).
8. The results of seven days of stock price volatility prediction are used for investment portfolio selection strategy.

RESULT AND DISCUSSION

The data used is Close Price data from several stock listed on the Indonesia Stock Exchange (IDX) starting from the period 2023-05-11 to 2024-05-08 which is shown in Figure 1. It can be seen that the last few months have experienced a very significant decrease in the daily value of shares, which is caused by the impact of the boycott.

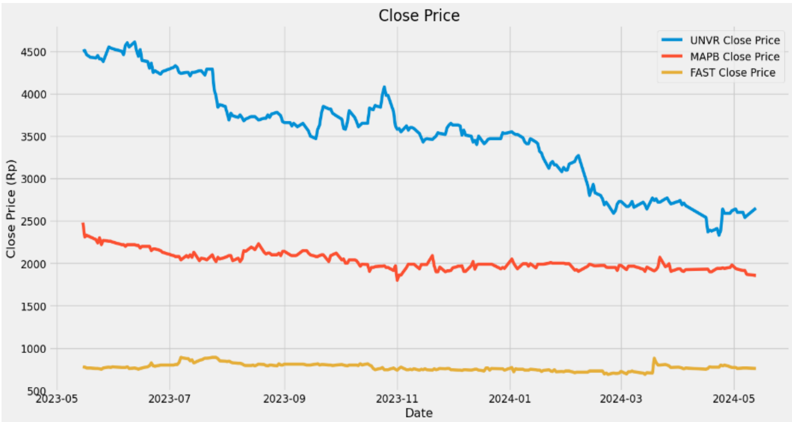

Figure 1. Graph Close Price of UNVR, MAPB, and FAST Shares

Based on the closing price data, the variance of return and log return can be calculated to determine the volatility of stock prices, from the three stocks the volatility is depicted in Figure 2. The graph shows that there are several extreme values, this shows volatility clustering which indicates that there is a serial dependence in the time series data. Volatility clustering indicates that the appropriate time series model for prediction is Long Short Term Memory (LSTM).

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-84, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

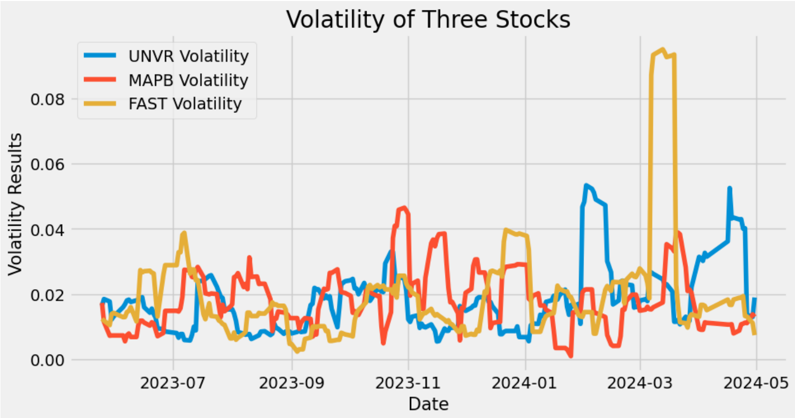

Figure 2. Graph of Stock Price Volatility Results on UNVR, MAPB, and FAST Shares

Prediction of stock return volatility is done by building the LSTM model which is done on Google Colaboratory platform with the programming language used is the Phyton 3 programming language. By using the parameters that have been described in the research methodology, the results of predicting stock price volatility in the next seven days with data divided into 80% training data and 20% testing data. With the evaluation of the model used, namely Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

Table 2. Model Evaluation

Stock Name	MSE	RMSE
UNVR	0.15397	0.15397
MAPB	0.17126	0.17126
FAST	0.26359	0.26359

Table 2 shows the results of model evaluation for the three stock are quite small, and it can be said that the model has good performance for predicting stock price volatility using LSTM.

Below are presented the volatility prediction results of the three stocks:

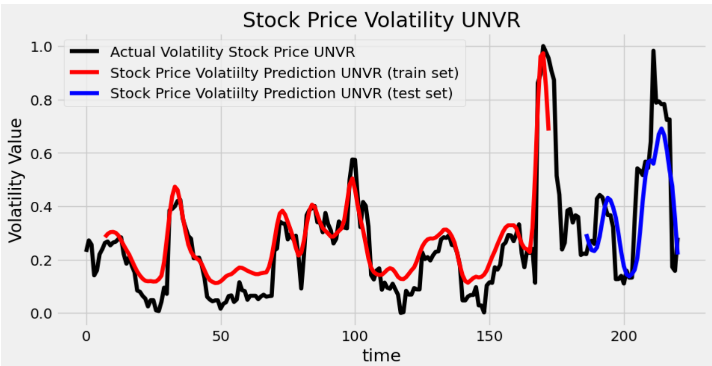

Figure 3. Stock Price Volatility UNVR

Figure 3 shows UNVR stock price volatility data along with stock price volatility prediction using LSTM. The black line represents the actual UNVR stock price volatility data. There are two other lines that represent the results of predicting UNVR stock price volatility data with LSTM, one for the training data is shown with a red line and a blue line to show the testing data. From Figure 3, it can be seen that in general, the prediction of stock price volatility with LSTM is able to follow the actual UNVR stock price volatility movement pattern quite well, although there are some differences.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-84, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Table 3. UNVR stock price volatility prediction

Day	Prediction Volatility
1	0.01592
2	0.01088
3	0.00763
4	0.00582
5	0.00543
6	0.00602
7	0.00720
Average	0.00841

Table 3 shows the predicted stock price volatility of UNVR for the next seven days that have been denormalized. Based on the prediction results, it decreased on day two to day five, and increased on day six and day seven, can be seen visualized with Figure 4.

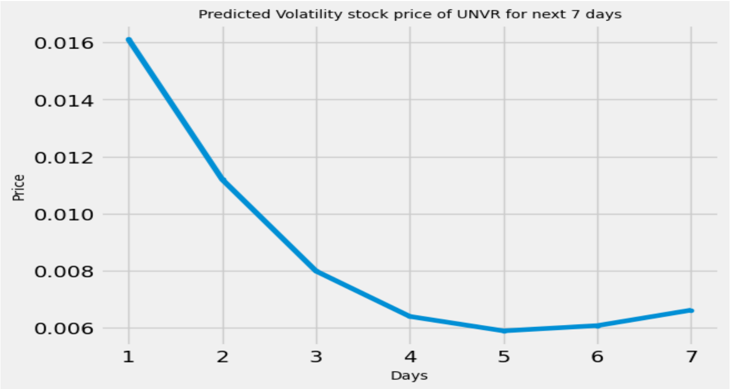

Figure 4. Predicted Volatility Stock Price of UNVR

Figure 5 shows the prediction of MAPB stock price volatility, showing that the training data is less accurate in following the actual stock price volatility movement pattern, and the test data shows a significant difference with the actual price, especially at the end of the graph. However, in general the prediction of stock price volatility using LSTM is able to follow the actual volatility movement pattern of MAPB stocks quite well.

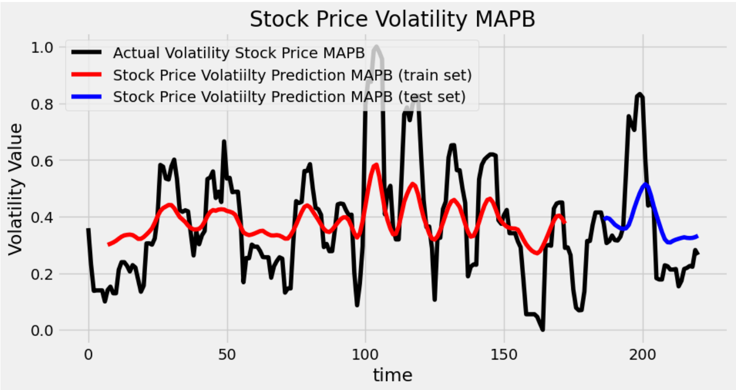

Figure 5. Stock Price Volatility MAPB

Table 4 shows the predicted stock price volatility of MAPB for the next days that has been denormalized. Based on the prediction results, the volatility of MAPB stock prices has increased from day one to day seven.

Table 4. MAPB stock price volatility prediction

Day	Prediction Volatility
1	0.01606
2	0.01642
3	0.01686
4	0.0173
5	0.01762
6	0.01787
7	0.01808
Average	0.01717

Table 4 will be visualized in Figure 6.

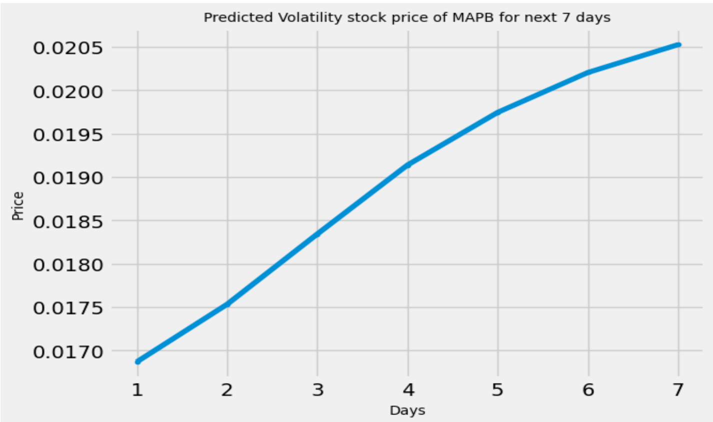

Figure 6. Predicted Volatility Stock Price of MAPB

Next, predict the volatility of the FAST stock price. Similar to Figure 3 and Figure 5, the volatility of UNVR and MAPB stocks is predicted. On FAST stock the LSTM model is generally able to follow the actual FAST stock price volatility movement pattern quite well.

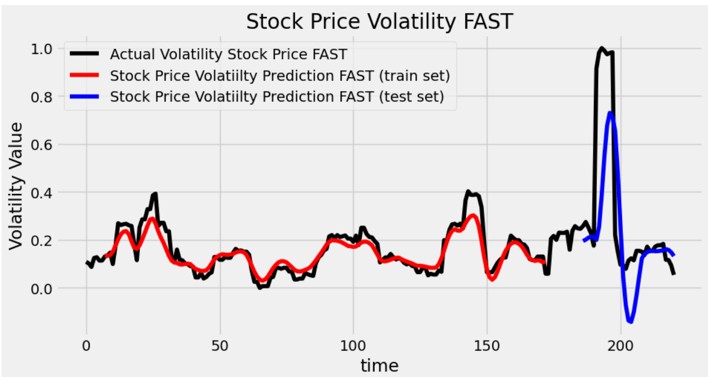

Figure 7. Stock Price Volatility FAST

The process of predicting stock price volatility for the next seven days on FAST stock, the predicted value decreased on day four and day five and increased again on days six and seven. As shown in Table 5 and Figure 8.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 05 May 2024
DOI: 10.47191/ijcsrr/V7-i5-84, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Table 5. FAST stock price volatility prediction

Day	Prediction Volatility
1	0.01465
2	0.01365
3	0.01296
4	0.01263
5	0.01266
6	0.01288
7	0.01318
Average	0.01323

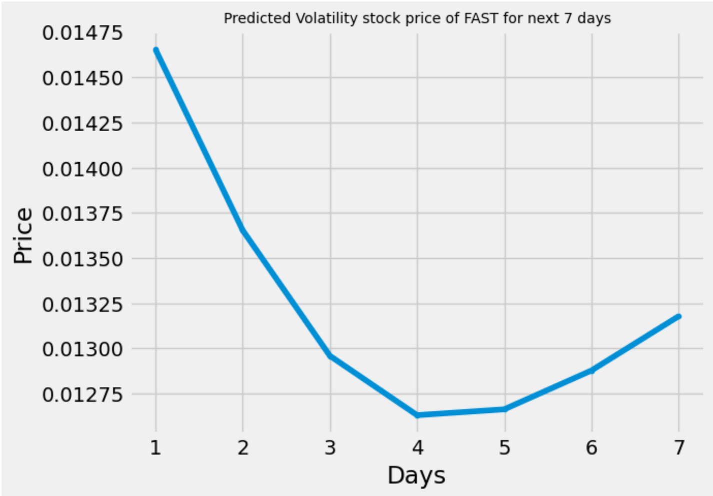

Figure 8. Predicted Volatility Stock Price of FAST

CONCLUSION

The Volatility prediction results of the three stocks for the following seven days were determined based on the results and talks that were held. It was discovered that the volatility of the stock prices of FAST and UNVR showed similar outcomes, namely experiencing a reduction followed by a rise. Meanwhile, the volatility of MAPB stock price has consistenly increased, and it can be concluded that by testing the model using MSE and RMSE produces low values, the volatility the three stocks can be modeled using LSTM. LTSM is able to store and remember historical data from short or long term. The low MSE and RMSE evaluation results indicate that the predicted value is almost the same as the actual data value, which indicates that the model can classify the movement of stock price volatility with a high level accuracy. This is what gives the results important linkage for an investor when making a more informed investment decision. The average return volatility of the three stocks shows the fact that in the Indonesian market that volatility often increases where there are bad news compared to changes in volatility during the period with no bad news for the sectoral stocks.

For the investment portfolio selection strategy, we place ourselves wanting to invest money into several stocks by looking at the volatility value of the available stock prices, resulting in low volatility in Unilever Indonesia shares, moderate volatility results in FAST shares and high volatility results in MAPB shares. So we can make an investment portfolio selection, if the expected time when investing is short in making an investment, we can invest in MAPB share and if we want to invest for a long time, we can invest in Unilever results and FAST because they have low volatility results.

Future research should be able to predict volatility using different models, or if it uses the same model, it should be able to alter the dataset separation, number of batches, hidden layer, and epochs, as well as add dropouts in the hidden layer. It is also anticipated that more data will be used, or different stocks from this study will be used, in an effort to improve the predictive models accuracy value. Additionally, in order to make more accurate and exact investment judgments, value indicators should be considered while choosing an investment portfolio.

REFERENCES

1. Z. Ouyang, M. Lu, dan Y. Lai, "Forecasting stock index return and volatility based on GAVMD- Carbon-BiLSTM: How important is carbon emission trading?," *Energy Econ.*, vol. 128, no. October, hal. 107134, 2023, doi: 10.1016/j.eneco.2023.107134.
2. Y. Yu, Y. Lin, X. Hou, dan X. Zhang, "Novel optimization approach for realized volatility forecast of stock price index based on deep reinforcement learning model," *Expert Syst. Appl.*, vol. 233, no. January, hal. 120880, 2023, doi: 10.1016/j.eswa.2023.120880.
3. Z. De Khoo, K. H. Ng, Y. B. Koh, dan K. H. Ng, "Forecasting volatility of stock indices: Improved GARCH-type models through combined weighted volatility measure and weighted volatility indicators," *North Am. J. Econ. Financ.*, vol. 71, no. January, 2024, doi: 10.1016/j.najef.2024.102112.
4. H. Markowitz, "Portofolio Selection," *J. Finance*, vol. 1, no. 7, hal. 77–91, 1952.
5. H. Y. Kim dan C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Syst. Appl.*, vol. 103, hal. 25–37, 2018, doi: 10.1016/j.eswa.2018.03.002.
6. M. F. Rizkilloh dan S. Widyanesti, "Cryptocurrency Price Prediction Using Long Short Term Memory (LSTM) Algorith," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, hal. 25–31, 2022, doi: 10.29207/resti.v6i1.3630.
7. R. Liu, Y. Jiang, dan J. Lin, "Forecasting the Volatility of Specific Risk for Stocks with LSTM," *Procedia Comput. Sci.*, vol. 202, hal. 111–114, 2022, doi: 10.1016/j.procs.2022.04.015.
8. M. Beniwal, A. Singh, dan N. Kumar, "Forecasting multistep daily stock prices for long-term investment decisions: A study of deep learning models on global indices," *Eng. Appl. Artif. Intell.*, vol. 129, no. June 2023, hal. 107617, 2024, doi: 10.1016/j.engappai.2023.107617.
9. A. Petrozziello *dkk.*, "Deep learning for volatility forecasting in asset management," *Soft Comput.*, vol. 26, no. 17, hal. 8553–8574, 2022, doi: 10.1007/s00500-022-07161-1.
10. A. Bash dan K. Alsaifi, "Fear from uncertainty: An event study of Khashoggi and stock market returns," *J. Behav. Exp. Financ.*, vol. 23, no. October 2018, hal. 54–58, 2019, doi: 10.1016/j.jbef.2019.05.004.
11. S. Shanaev dan B. Ghimire, "Is all politics local? Regional political risk in Russia and the panel of stock returns," *J. Behav. Exp. Financ.*, vol. 21, no. September 2017, hal. 70–82, 2019, doi: 10.1016/j.jbef.2018.11.002.
12. Y. Xu, T. Liu, dan P. Du, "Volatility forecasting of crude oil futures based on Bi-LSTM-Attention model: The dynamic role of the COVID-19 pandemic and the Russian-Ukrainian conflict," *Resour. Policy*, vol. 88, no. November 2023, hal. 104319, 2024, doi: 10.1016/j.resourpol.2023.104319.
13. Y. Maulana, "Modeling The Volatility of Sectoral Stock Price Indices in Indonesia," *Log. J. Penelit. Univ. Kuningan*, vol. 13, no. 1, hal. 53–72, 2022.
14. E. Chotib dan N. Huda, "Analysis of the Effect of Macroeconomic Variables On the Indonesian Sharia Stock Index (ISSI) for the Period 2016 - 2019," *JEBA (Journal Econ. Bus. Aseanomics)*, vol. 4, no. 1, hal. 174–195, 2020, doi: 10.33476/j.e.b.a.v4i1.1235.
15. A. Atkins, M. Niranjan, dan E. Gerding, "Financial news predicts stock market volatility better than close price," *J. Financ. Data Sci.*, vol. 4, no. 2, hal. 120–137, 2018, doi: 10.1016/j.jfds.2018.02.002.
16. A. García-Medina dan E. Aguayo-Moreno, "LSTM–GARCH Hybrid Model for the Prediction of Volatility in Cryptocurrency Portfolios," *Comput. Econ.*, 2023, doi: 10.1007/s10614-023-10373-8.
17. J. Schmidhuber, "Lstm Can Solve Hard Long Time Lag Problems," vol. 9, hal. 473–479, 1997.
18. T. Roncalli, *Introduction to risk parity and budgeting*. New York: CRC Press, 2013.
19. D. Syafitri dan C. Batubara, "Evaluation of Islamic Banks Investment Strategies in the Face Of Financial Market Volatility," *J. Ilmu Komputer, Ekon. dan Manaj.*, vol. 3, no. 2, hal. 3565–3566, 2023.

Cite this Article: Ana Khofifahturizqi, Farikhin, Ratna Herdiana, Yola Ulrike Sihombing (2024). Prediction of Stock Price Volatility Using the Long Short Term Memory (LSTM) Model for Investment Portfolio Selection Strategy. International Journal of Current Science Research and Review, 7(5), 3259-3267