International Collaborative Conference on Multidisciplinary Science
Volume. 1, Number. 1, Year 2024
Page 86-96
Available online at: https://iccms.ifrel.org/index.php/ICCMS

Evaluation of Adam, Nadam, RMSProp and SGD Optimization Algorithms on LSTM Model for JISDOR Exchange Rate Prediction

Ariska Fitriyana Ningrum
Universitas Muhammadiyah Semarang, Indonesia
Corresponding author: ariskafitriyana@unimus.ac.id

Abstract: *This research explores the application of four optimization algorithms—Adam, Nadam, RMSProp, and SGD—on a Long Short-Term Memory (LSTM) model to forecast the Jakarta Interbank Spot Dollar Rate (JISDOR). The volatile nature of exchange rate data, influenced by global and domestic economic dynamics, necessitates the use of models like LSTM that excel in capturing both short- and long-term dependencies. Performance was assessed using metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). Among the optimizers, Nadam proved to be the most effective, achieving the lowest RMSE of 62.767 and a MAPE of 0.003, indicating its capability in managing complex fluctuations in the dataset. Despite Nadam's promising results, opportunities for improvement remain, including the inclusion of additional input variables, fine-tuning model parameters, and expanding the training dataset. This study underscores the critical role of selecting appropriate optimization algorithms for enhancing the accuracy of LSTM models in forecasting volatile financial time-series data, particularly for currency exchange rates.*

Keywords: *Jakarta Interbank Spot Dollar Rate (JISDOR), LSTM, Optimization Algorithm, Time Series.*

1. INTRODUCTION

The prediction of exchange rates, particularly for the Jakarta Interbank Spot Dollar Rate (JISDOR), plays a pivotal role in financial forecasting, as fluctuations in exchange rates can significantly influence economic policies, market stability, and investment decisions (Chong , Han and Park 2017). Time series forecasting has been widely used to model exchange rate behavior, given its inherent temporal dependencies and volatility (Sule, A., and Ozturk, H 2022). Among various time series models, the Long Short-Term Memory (LSTM) network has emerged as a powerful tool for handling such tasks due to its ability to capture long-term dependencies in sequential data (Hochreiter, S., and Schmidhuber, J. 1997). LSTM's robust performance in tasks involving non-linear dynamics, like exchange rate prediction, is a key feature that sets it apart from traditional statistical methods (Zahra, Saadah and Rismala 2021)

LSTM models rely on optimization algorithms to tune the weights and parameters during training. These optimization techniques determine the convergence speed and the efficiency of model training, which directly affects prediction accuracy (Kingma and Ba 2015). Optimization algorithms like Adam (Adaptive Moment Estimation), Nadam (Nesterov-accelerated Adaptive Moment Estimation), RMSprop (Root Mean Square Propagation), and SGD (Stochastic Gradient Descent) are popular choices in deep learning because they adapt the learning rates during training to improve model performance (Reddi, S. J, Kale, S and Kumar, S. . 2018). Adam, in particular, has proven effective due to its combination of momentum and adaptive learning rate, which accelerates convergence and mitigates challenges

Received: April 30, 2024; Revised: May 30, 2024; Accepted: June 18,2024; Published: June 30, 2024

such as vanishing gradients (Kingma & Ba, 2015). Nadam builds on Adam by incorporating Nesterov momentum, offering potential improvements in optimization efficiency (Dozat. T 2016)

The JISDOR exchange rate, as a time series dataset, presents a significant challenge for modeling due to its volatility, which is influenced by a complex interplay of domestic and international economic events (Bekaert, and Hodrick, 2019). A graphical representation of JISDOR from November 2024 reveals marked fluctuations, with periods of rapid depreciation followed by moments of relative stability, reflecting shifts in market sentiment, fiscal policies, and global trade dynamics (Sule, A and Ozturk, H. 2022). For instance, sudden drops or surges in the JISDOR rate may be linked to global economic events or policy changes in Indonesia, such as changes in interest rates or inflation expectations. The volatility seen in the JISDOR time series requires a robust model that can adapt to these shifts and provide accurate forecasts.

Figure 1 illustrates the JISDOR exchange rate from November 2024, displaying notable volatility due to both internal and external economic factors. This chart serves as a testament to the dynamic nature of currency markets, where rates are often subject to rapid fluctuations influenced by a wide array of macroeconomic factors such as inflation rates, political stability, and international trade relations (Cai, Joo and Zhang 2020). Such variability underscores the need for sophisticated prediction models that can manage these complexities and yield accurate forecasts, especially for financial institutions and policymakers (Hyndman, R. J and Athanasopoulos, G 2018)

This study seeks to evaluate the effectiveness of different optimization algorithms within the LSTM model for predicting the JISDOR exchange rate. Optimization algorithms such as Adam, Nadam, RMSprop, and SGD have distinct advantages depending on the dataset characteristics and problem complexity. For example, Adam and Nadam are well-regarded for their adaptive learning rates and ability to handle noisy data, while RMSprop adjusts the learning rate based on the moving average of recent gradients. On the other hand, while SGD is a foundational method, it is often slower to converge and less effective in handling large datasets with complex patterns (Lecun, Y., et al. 2012). By comparing these algorithms, we aim to identify the most suitable optimizer for JISDOR exchange rate prediction.

The effectiveness of these optimization algorithms is assessed using performance metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), which provide insights into the accuracy and reliability of the predictions (Hyndman, R. J and Athanasopoulos, G 2018). RMSE is widely used in regression tasks, as it penalizes larger errors, while MAPE offers a more interpretable metric by providing a percentage error relative to actual values. This comparative analysis will help determine which optimizer delivers the most accurate prediction for the JISDOR exchange rate, contributing to the body of knowledge on time series forecasting and optimization in financial applications.

In conclusion, this research explores the optimization of LSTM models using four popular algorithms—Adam, Nadam, RMSprop, and SGD—to forecast JISDOR exchange rates. The outcomes of this study will shed light on the capabilities of these algorithms in managing the complexities of financial time series and offer practical insights for enhancing forecasting models in the volatile domain of currency exchange rates.

2. LITERATURE REVIEW

Long Short Term Memory

Long short-term memory (LSTM) is one of the methods in neural networks which was first discovered by Sepp Hochreiter and Jurgen Schmiduber in 1997. Long short-term memory is the development of the repetitive neural network method. (RNN), this is because there are many weaknesses in the RNN method, namely long-term delays that cannot be accessed in an architecture (Greff, et al. 2017).

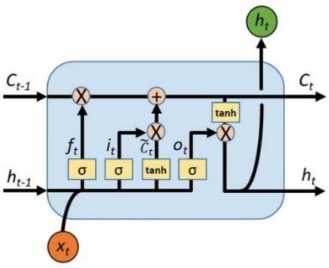

Figure 1. LSTM Architecture

Forecasting is done using LSTM. The LSTM architecture consists of a memory cell and three gates: input gate (i_t)forget gate (f_t), and output gate (O_t). The input gate plays a role in regulating how much information must be stored in a sold state. The gate prevents the cell from storing useless data. Forget gate is responsible for setting a fixed value in the memory cell. Output Gate regulates how many values in the memory cell will be used for output. There are several computational stages in the LSTM method, which can be seen in. The calculation of the forget gate value by combining the input variable and the hidden state value in period $t - 1$ (h_{t-1}) becomes the following equation.

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad (4)$$

Next, the input gate value and candidate cell state are calculated using a combination of the input variable value and the previously hidden state value at t-1 with the aim of updating the cell state. The search for the input gate value involves the sigmoid activation function and getting the candidate cell state value using the hyperbolic tangent activation function.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{5}$$

$$\tilde{c}_t = tanh(W_c x_t + U_c h_{t-1} + b_c), \tag{6}$$

After getting the values for the forget gate, input gate, and candidate cell state, then proceed with creating a new cell state (C_t) in a neuron which is a combination of the resulting value from equation (1) which is multiplied by the previous cell state $(C_{t-1})$is then added with the product of the result obtained in equation (4) and the result of equation (5). Here are the similarities

$$C_t = i_t \times \tilde{C}_t + f_t \times C_{t-1} \tag{7}$$

The value at the output gate can be searched using a combination of the previous input and hidden state and involves the sigmoid activation function which can be seen in the following equation.

$$O_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{8}$$

$$h_t = O_t \times tanh\ (C_t). \tag{9}$$

$f_t, i_t, \tilde{c}_t, O_t$ successively the forget gate, input gate, intermediate cell state, and output gate. σ and tanh as the activation function. W_f, W_i, W_c, W_o is the weight value, $b_f, b_i, b_c, b_o$is the bias value. h_{t-1} is the previous period input value and x_t input value on t. c_{t-1} is the old state, c_t is the current cell state, h_t is the hidden state.

Model Evaluation with Optimization Algorithm

After the prediction model is successfully built, the training process will be carried out until it reaches the desired level of accuracy. At this stage, various optimization algorithms such as Stochastic Gradient Descent (SGD), Root Mean Square Propagation (RMSProp), Adaptive Moment (Adam), and Nesterov Adam (Nadam) will be used. Stochastic Gradient Descent (SGD) is one of the variants of the gradient descent optimization algorithm that performs parameter updates for each trained data. In the parameter update process, SGD does not perform repeated iterations, making it more efficient for large datasets (S. Ruder 2016). Meanwhile, the performance and objectives of the RMSProp algorithm are very similar to Adadelta, but RMSProp is superior to AdaGrad in reducing the learning rate. Adaptive Moment Estimation (Adam) is an optimization method that calculates the learning rate adaptively for each parameter. Like Adadelta and RMSProp, Adam also stores the average gradient of the previous process exponentially. While Nadam (Nesterov-accelerated Adaptive Moment Estimation) is a combination of RMSProp and Nesterov Accelerated Gradient (NAG).

Calculating Forecasting Errors

Hyndman & Koehler (2006) stated that RMSE can be used to compare different methods on data with the same scale. (Makridakis, Wheelwright and McGee 1999). Therefore, the evaluation of the model use RMSE with the following equation.

Therefore, the evaluation of the model use RMSE with the following equation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{T}(y_t - \hat{y}_t)^2}{T}} \quad (10)$$

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_t - \hat{y}_t}{T}\right|\right) \times 100 \quad (11)$$

Definition:

t : time-*t*

y_t : actual data ke-*t*,

$\hat{y}_t$: forecast data ke-*t*

T : number of out sample data.

3. METHODS

Data and Source

The data used in this study is secondary data obtained from Bank Indonesia's public data set regarding the Jakarta Interbank Spot Dollar Rate (JISDOR). The data includes the value of the exchange rate from May 2013 to November 2024. The following is the structure of the time series data used, available in the following table.

Table 1. Research Data Structure

Date	Y_t
20/05/2013	y_1
20/05/2013	y_2
20/05/2013	y_3
⋮	⋮
⋮	⋮
30/11/2024	y_n

Research Procedure

This research starts by collecting historical data of currency exchange rates from reliable sources, then cleaning the data by removing missing values and normalizing the data to ensure uniformity. After that, the data is divided into 80% for training and 20% for testing. The LSTM model is constructed by determining the appropriate number of layers and units in each hidden layer, and adding a Dense layer as the output layer to predict future exchange rates. Furthermore, three optimization algorithms, namely Stochastic Gradient Descent (SGD), Adaptive Moment (Adam), and Nesterov Adam (Nadam), were selected to be used in model training. The training parameters, such as epoch, batch size, and learning rate, were determined, and the model was trained with training data and evaluated with testing data. Model performance is evaluated using two metrics, namely Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

4. RESULTS

JISDOR exchange rate prediction is one of the challenges in time series analysis due to data fluctuations influenced by global and domestic economic factors. In this study, the visualization of time series data shows varying volatility patterns, with sharp fluctuations in certain periods that require a reliable prediction model approach. The LSTM model, which is known to excel in capturing long-term dependency patterns, was tested using three optimization algorithms-Adam, RMSprop, and SGD. The following illustrates the fluctuation of the time series plot on the JISDOR exchange rate.

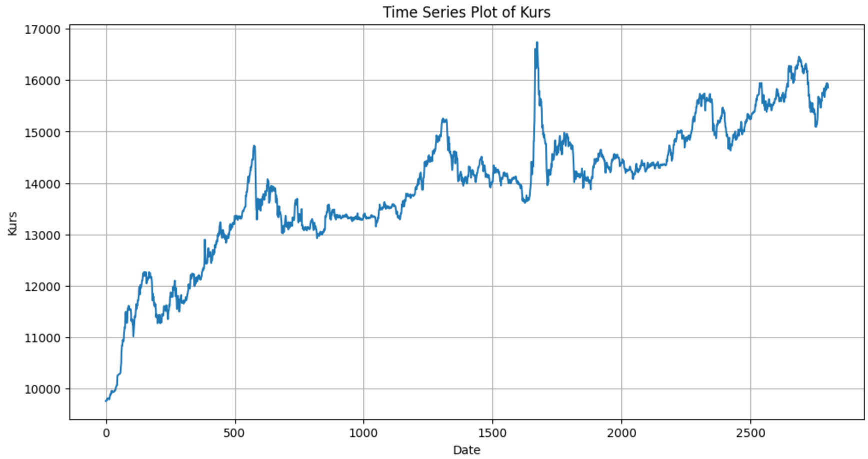

Figure 2. Time Series plot of JISDOR from 2013 - 2024

The time series chart of the JISDOR exchange rate from 2013 to 2014 shows an upward trend in the exchange rate, reflecting the depreciation of the domestic currency against the USD during the period. Significant fluctuations are seen at some points, indicating high market volatility, possibly due to changes in monetary policy or the impact of global economic events. Nonetheless, there are certain intervals where the exchange rate is relatively stable, reflecting calmer market conditions. Overall, this chart reflects the dynamics of the foreign exchange market which is influenced by various internal and external factors.

Furthermore, the construction of the prediction model is carried out using the LSTM model. However, before that, the initial steps before the LSTM modeling process include data normalization. At this stage using MinMaxScaler to normalize the data to a range of 0 to 1.

Table 2. Data Normalization Result

Date	Y_t
20/05/2013	0,000716
20/05/2013	0,000716
20/05/2013	0,002005
⋮	⋮
⋮	⋮
30/11/2024	0,873227

This normalization is done so that all values are on the same scale, which is important to improve the performance of the LSTM model as LSTM is sensitive to large values in the data. Next, the data is split into two parts. In this study, 80% training data and 20% testing data were used. This separation aims to train the model on the training data and evaluate the model performance on the testing data. Selanjutnya, dataset diubah menjadi format yang sesuai untuk LSTM dengan menggunakan fungsi **create_dataset**, yang membentuk pasangan input-output berdasarkan parameter **look_back**, yaitu jumlah langkah waktu sebelumnya yang digunakan untuk memprediksi langkah waktu berikutnya. Setelah itu, data input diubah dimensinya menjadi tiga dimensi [**samples, time steps, features**] agar kompatibel dengan input LSTM. Proses ini mempersiapkan data secara optimal sebelum memasuki tahap pemodelan.

In this study, a comparison of RMSE and MAPE values between 3 optimization algorithms will be carried out after the training and testing process..

Table 3. Accuracy Value in Optimization Algorithms

Optimizatioin Type	RMSE	MAPE
Stochastic Gradient Descent (sgd)	178.437	0.010
Adaptive moment (Adam)	66.860	0.003
Root Mean Square Propagation (RMSProp)	145.062	0.008
Nesterov Adam (Nadam)	62.767	0.003

Based on the performance evaluation results of LSTM models with four types of optimization (SGD, Adam, RMSprop, and Nadam), Nadam shows the best performance with an RMSE of 62,767 and a MAPE of 0.003, followed by Adam with an RMSE of 66,860 and a MAPE of 0.003, reflecting the ability of these two methods to set adaptive learning rates and accelerate model convergence. Meanwhile, RMSprop has an RMSE of 145.062 and a MAPE of 0.008, showing a fairly good performance but still inferior to Adam and Nadam. On the other hand, SGD has the worst performance with an RMSE of 178,437 and MAPE of 0.010, due to its slow convergence and difficulty in reaching the global minimum. Overall, the Nadam and Adam optimization methods are superior in minimizing the prediction error, making them more effective in the LSTM model for this data.

Some of the time series plots below illustrate the comparison of predicted results and actual data in each optimization algorithm.

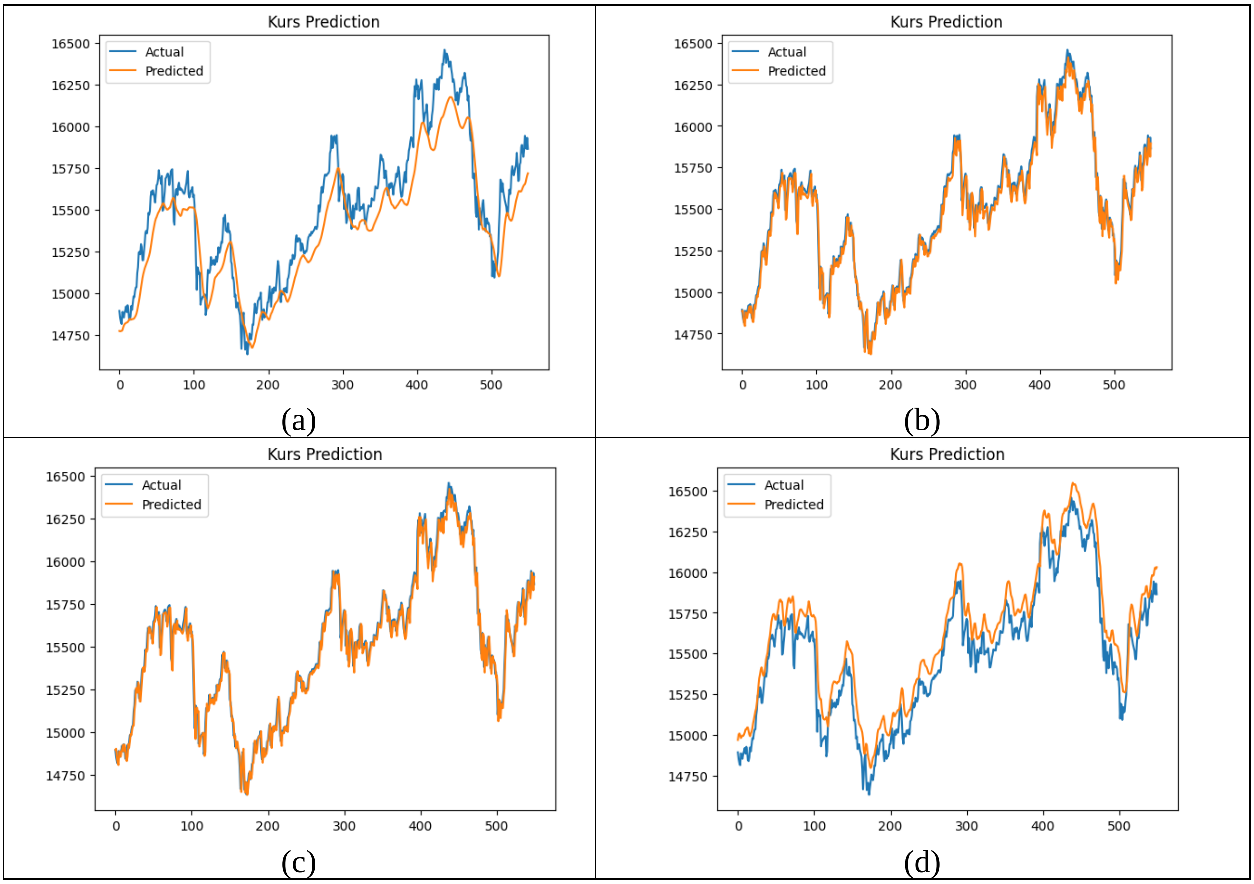

Figure 3. Kurs Prediction Using (a) SGD (b) ADAM (c) Nadam (d) RMSProp optimization algorithm

Rate prediction using several optimization algorithms shows that the prediction data pattern using the Nadam and Adam Algorithms has a pattern that is almost the same as the actual data pattern. So in this study, it was found that the best optimization for forecasting JISDOR values in LSTM modeling is to use Nadam Optimization.

5. DISCUSSION

The prediction of the JISDOR exchange rate is a complex challenge in time series analysis due to its volatility, influenced by both global and domestic economic factors. The time series plot reveals periods of high fluctuation, which can be attributed to market turbulence, possibly driven by changes in monetary policy or global economic shifts. These fluctuations underline the necessity for a robust predictive model capable of capturing long-term dependencies. The LSTM model, known for its ability to capture such patterns, was evaluated using three optimization algorithms: SGD, Adam, RMSprop, and Nadam. From the evaluation, it is clear that both Nadam and Adam optimization methods outperformed the other algorithms. With RMSE values of 62.767 and 66.860, respectively, and MAPE values of 0.003, both methods demonstrated their effectiveness in minimizing prediction errors by adjusting learning rates dynamically. In contrast, RMSprop performed reasonably well with an RMSE of 145.062 and a MAPE of 0.008 but still fell short compared to Adam and Nadam. SGD, on the other hand, performed the worst, with an RMSE of 178.437 and a MAPE of 0.010, likely due to its slower convergence rate and difficulty in reaching the global minimum. The results clearly indicate that Nadam and Adam provide the most reliable predictions, making them more suitable choices for LSTM model optimization in this context.

Nadam and Adam are more effective compared to other optimizations such as RMSprop and SGD in JISDOR exchange rate prediction because they combine the advantages of momentum techniques and adaptive learning rate adjustment. Adam, developed by Kingma and Ba (2014), combines momentum and RMSprop by storing the first and second gradient estimates, enabling more efficient and stable learning rate adjustment in large and volatile data. Nadam, which is a variant of Adam with the addition of Nesterov Accelerated Gradient (NAG), further improves convergence by predicting the gradient before parameter updates, as described by Dozat (2016). On the other hand, RMSprop is more limited as it only adjusts the learning rate based on the mean square of the gradient, and SGD, although simple, tends to be slower in achieving convergence on highly volatile data. Therefore, Adam and Nadam are superior in handling volatile time series data such as JISDOR exchange rates, which are affected by various external and internal factors that cause sharp fluctuations.

6. CONCLUSION

The use of LSTM (Long Short-Term Memory) method in time-series data-based prediction is very effective because of its ability to remember and store information from data in the short and long term. This research applies LSTM to non-linear variables in the form of exchange rates that tend to fluctuate every day. Based on the results of the prediction model compared using four optimization algorithms, Nesterov Adam (Nadam) showed the best accuracy, with the lowest RMSE value of 62.767 and MAPE of 0.003. Nevertheless, the accuracy obtained is still far from ideal expectations. Several factors that can be evaluated in future research, such as the addition of input variables, variations in the number of epochs, number of hidden layers, batch size, as well as the amount of training and testing data, can potentially improve the accuracy of this prediction model.

LIMITATION

The limitations of this study are mainly related to the limitations of the model architecture and the data used. Although Nesterov Adam (Nadam) optimization shows the best performance with the lowest RMSE and MAPE, the accuracy obtained is still far from ideal expectations. Some factors that may affect this result include the selection of input variables that may not fully cover all factors affecting the JISDOR exchange rate. In addition, the performance of the model can be improved by experimenting with various hyperparameters such as the number of epochs, number of hidden layers, batch size, and proportion of training and testing data. This research is also limited to the use of only four optimization algorithms, so the incorporation of more advanced or hybrid optimization techniques may yield better results. Future research can explore these aspects to improve the accuracy and robustness of the prediction model**.**

REFERENCES

Bekaert,, Geert , dan Robert Hodrick,. 2019. International Financial Management. Pearson Education. doi: http://dx.doi.org/10.1017/9781316282274.

Cai, Fang, Hyunsoo Joo, dan Zhiwei Zhang. 2020. “The impact of global economic events on currency exchange rates: Evidence from China.” Journal of International Economics. doi: http://dx.doi.org/10.2139/ssrn.1434412.

Chong , Eunsuk, Chulwoo Han, dan Franck. C Park. 2017. “Deep Learning Network for Stock Market Analysis and Prediction: Methodology, Data Representations and Case Studies.” Expert Systems with Applications 83: 187 - 205. doi: https://doi.org/10.1016/j.eswa.2017.04.030.

Dozat. T. 2016. “Dozat, T. (2016). Incorporating Nesterov momentum into Adam. ICLR Workshop.” Proceedings of the 4th International Conference on Learning Representations 11. doi: https://doi.org/10.4236/blr.2019.104046.

Greff, Klaus , Rupesh K Srivastava, Jan Koutnik, Bas R Steunebrink, dan Jurgen Schmidhuber. 2017. “LSTM: A Search Space Odyssey.” IEEE Transactions on Neural Networks and Learning Systems 28: 2222-2232. doi: https://doi.org/10.1109/TNNLS.2016.2582924.

Hochreiter, S.,, dan Schmidhuber, J. 1997. “Long short-term memory. Neural Computation.” 9: 1735–1780. doi:doi:10.1162/neco.1997.9.8.1735.

Hyndman, R. J , dan Athanasopoulos, G. 2018. Forecasting: principles and practice. OTexts.

Kingma, Diederik P, dan Jimmy Lei Ba. 2015. “Adam: A Method for Stochastic Optimization.” Published as a conference paper at the 3rd International Conference for Learning Representations. doi: https://doi.org/10.48550/arXiv.1412.6980.

Lecun, Y., Bottou, L., , Orr, G. B., dan Müller, K.-R. 2012. Efficient backprop. In Neural Networks: Tricks of the Trade. Springer.

Makridakis, S, S.C Wheelwright, dan V.E McGee. 1999. Metode Aplikasi Peramalan. Jakarta: Erlangga.

Reddi, S. J, Kale, S, dan Kumar, S. . . 2018. "On the convergence of Adam and beyond." International Conference on Learning Representations (ICLR). doi: https://doi.org/10.48550/arXiv.1904.09237.

S. Ruder. 2016. An overview of gradient descent optimization. https://arxiv.org/abs/1609.047477.

Sule, A, dan Ozturk, H. . 2022. "The impact of global financial crises on exchange rate volatility." Journal of Financial Economics, 35.

Sule, A., , dan Ozturk, H. 2022. "The impact of global financial crises on exchange rate volatility. Journal of Financial Economics." Journal Of Financial Economics 102-112. doi: http://dx.doi.org/10.35527/kfedoi.2015.14.4.004.

Zahra, Hasna Haifa, Siti Saadah, dan Rita Rismala. 2021. "The Foreign Exchange Rate Prediction Using Long Short Term Memory: A Case Study in Covid-19 Pandemic." International Journal on Information and Communication Technology 6: 94-105. doi: http://dx.doi.org/10.21108/IJOICT.2020.62.538.