Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

Performance Comparison of LSTM Models with Various Optimizers and Activation Functions for Garlic Bulb Price Prediction Using Deep Learning

Dasril Aldo*[1], Adanti Wido Paramadini[2], Muhammad Afrizal Amrustian[3]

[1]Department of Informatics, Telkom University, Indonesia
[2]Biomedical Engineering, Telkom University, Indonesia
[3]Department of Information and Computer Science, King Fahd University of Petroleum and Minerals, Saudi Arabia

Email: [1]dasrilaldo@ telkomuniversity.ac.id

Received : Feb 5, 2025; Revised : Mar 5, 2025; Accepted : Mar 6, 2025; Published : Apr 26, 2025

Abstract

Accurate commodity price forecasting is crucial for market stability and decision-making. This study evaluates the performance of the Long Short-Term Memory (LSTM) model using various activation functions and optimization algorithms for predicting garlic bulb prices. Historical price data was collected from panelharga.badanpangan.go.id and preprocessed through normalization and dataset splitting into training, validation, and test sets. The model was trained for 200 epochs using activation functions ReLU, Sigmoid, and Tanh, combined with optimization algorithms Adam, RMSprop, SGD, Adagrad, Adadelta, Nadam, and AdamW. Experimental results indicate that ReLU + Adam achieves the best performance with Final Epoch Loss of 0.001789, RMSE of 0.701632, MAPE of 0.009593, and R^2 of 0.909794, followed by Sigmoid + Nadam and Tanh + Adam, which also yielded high accuracy. These findings reinforce prior research, highlighting Adam and its momentum-based variants as effective optimizers for LSTM training. This study provides insights into selecting optimal activation functions and optimizers for commodity price forecasting. Future work may explore hybrid models and external factors, such as global market trends, to enhance predictive accuracy in time series data analysis.

Keywords : *Activation Function, Commodity Price Prediction, LSTM, Optimization Algorithm, Time Series Analysis.*

1. INTRODUCTION

Price prediction is an element of knowledge-based and strategic decision-making in finance and commodities. Having an idea of what is going to happen helps companies and investors mitigate risks and increase profits. Long Short-Term Memory (LSTM), a recurrent neural network (RNN) type, has emerged as the most popular way to handle complex time series data thanks to its capacity of addressing gradient vanishing problems and long term keeping [1], [2].

The LSTMs have a cell state and gate mechanism that provide the ability for this model to retain critical bits of information for long time intervals while other traditional RNN models cannot [3], [4]. Although LSTMs excel in processing time-series data, their performance is highly sensitive to hyperparameter combinations (e.g., activation functions and optimizers) [5], [6]. According to research, when we choose activation functions like ReLU, Tanh and Linear or Optimizers like Adam, RMSProp or Nadam it may have different results effect on model performance [7], [8]. Other studies have compared LSTMs to models such as ARIMA and Support Vector Machine (SVM) (Kar v. Despite the fact that while ARIMA is better suited for short term predictions, it also infrequently accompanies LSTMs as they are able to work better with non-linear and complex data for distant futures [9]. In

Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

addition, pure LSTMs can to some extent be improved in prediction accuracy via hybrid models such as DQN-LSTM [10]. Furthermore, [11], [12] find that Optimized LSTM models using other algorithms such as Glowworm Swarm Optimization (GSO), and LightGBM can enhance prediction performance over conventional methods respectively.

According to the studies conducted on [13] Tanh activation function is better for some situations which yield high accuracy but it was found that ReLU is more suitable option when trying to reduce loss expects downsides of having high amount of noise. Similarly, LSTM has also been shown to perform better than GRU and RNN in the context of long-term data dependence [14], [15]. Conversely, studies on stock price predictions in different kinds of markets indicate that the selection of activation functions can drastically impact accuracy [16]. According to the research done by [17], prediction accuracy can be enhanced further by combining LSTM with regression-based decision making strategies and dynamic programs. In studies like [18], hybrid methods like SSA-LSTM were used to increase prediction accuracy by filtering the noise in time-series data. Similarly, LSTM models tuned by genetic algorithms can make much more accurate predictions than other prediction models as well [19].

Despite this tremendous achievement, literature on how to best pair activation functions and optimizers that leads to LSTM performance maximization is missing. The objective of this study is to investigate and determine the combination of activation and optimizer functions to increase LSTM model prediction accuracy for financial and commodity time series data. With further, deeper exploration, these findings will likely play an important role in enhancing prediction accuracy and better informing business decision-making [1], [6], [20].

2. METHOD

In this study, garlic bulb price prediction was modeled using Long Short-Term Memory (LSTM) model and its effectiveness was evaluated. Abstract The study is concerned with testing the various activation functions and optimizing algorithms combinations in order to compare them to achieve best predictions. It consists of different stages such as data collection, data preprocessing,building the LSTM model and evaluating the model.

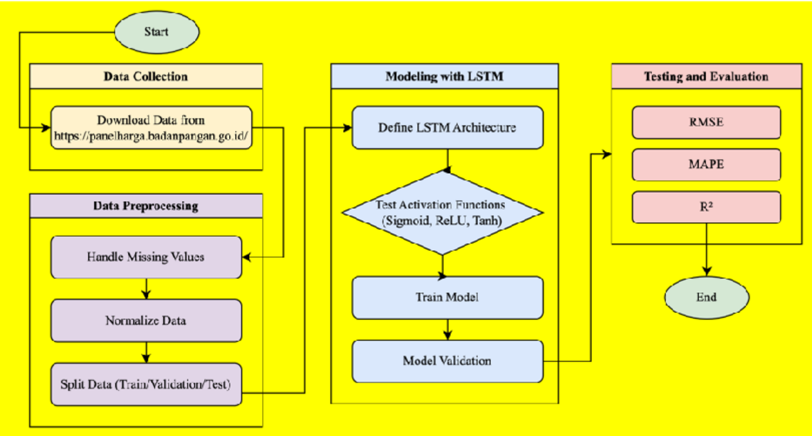

Figure 1. Research Methodology

Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

The research methodology consists of several key stages, as illustrated in Figure 1. This process includes data collection, preprocessing, model training, and evaluation to ensure optimal performance. Here is an explanation for each stage:

1. Data Collection: Historical data on Garlic Bulb prices is taken from reliable sources https://panelharga.badanpangan.go.id/ This data contains information that covers a long enough period of time to assist the model in capturing complex patterns and trends.

Table 1. Data Garlic Bulb

No	Date	Price
1	15/03/2021	35
2	16/03/2021	36.92
3	17/03/2021	35.25
4	18/03/2021	36.78
5	19/03/2021	36.78
6	20/03/2021	36.86
7	21/03/2021	-
8	22/03/2021	37.86
9	23/03/2021	37.19
10	24/03/2021	36.32
11	25/03/2021	38.42
12	26/03/2021	36.11
13	27/03/2021	36.91
14	15/03/2021	35
…	…	…
…	…	…
1333	06/11/2024	51.48
1334	07/11/2024	51.91

The collected data is summarized in Table 1, which presents historical garlic bulb price records. This data serves as the foundation for model training, enabling the LSTM model to learn price patterns over time. Next, the data will be visualized in graphical form to illustrate daily price trends. The resulting chart will depict price fluctuations, helping to identify significant changes in garlic bulb prices, as shown in Figure 2.

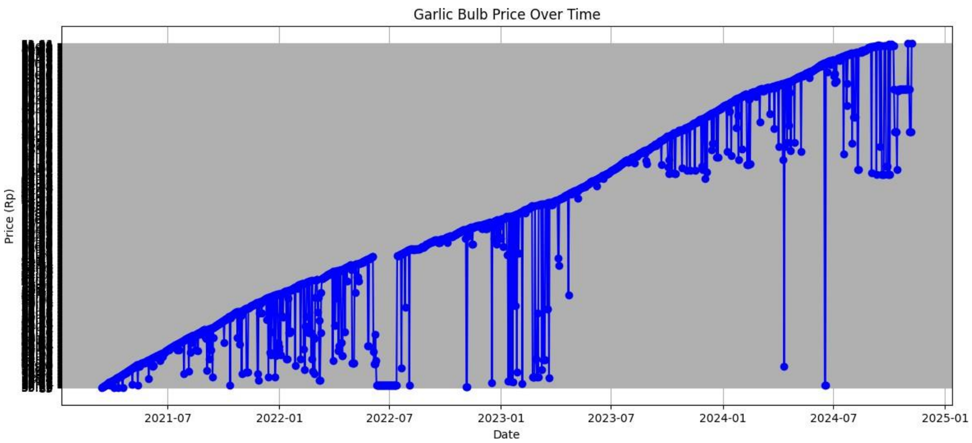

Figure 2. Commodity Price Trend Chart

Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

Figure 2 illustrates the price trend of garlic bulbs over time, showing fluctuations in commodity prices. The chart highlights both gradual increases and sharp decreases, providing insights into market volatility and seasonal price variations.

2. Praprocessing Data:
   a. Handling Missing Values: Each gap in the data is filled in using a linear interpolation method to maintain data integrity so that the training process is not interrupted by empty values.
   b. Data Normalization: Data is converted to a scale of [0.1][0, 1][0.1] using MinMaxScaler to facilitate model training and improve its stability. Normalization formula [21]:

$$X' = \frac{(X - X_{min})}{X_{max} - X_{min}} \quad (2)$$

   where is the original value, where ′ is the value after normalizationxx.
   c. Data Sharing: The processed data is divided into training (70%), validation (15%), and testing (15%) data. This is done to ensure the model is trained, validated, and tested with different data. In this study, the dataset was divided into 70% training, 15% validation, and 15% testing sets. This 70-15-15 split is commonly used in time series forecasting and deep learning applications to ensure a balanced distribution of data for model training and evaluation. The 70% training set allows the LSTM model to learn long-term patterns and trends from historical price data effectively. The 15% validation set is used to fine-tune hyperparameters and prevent overfitting by monitoring the model's performance on unseen data. The 15% testing set provides an unbiased evaluation of the final trained model's predictive accuracy on completely new data.

3. Modeling with LSTM:
   LSTM Model Architecture: The LSTM model is designed with a single LSTM layer containing 50 units, followed by a dense layer to predict a single output. This architecture is designed to address the problem of vanishing gradients and handle long-term dependencies in data. Activation Function[22]:
   a. ReLU (Rectified Linear Unit): This function returns the output . ReLU is used because of its efficiency in overcoming gradient vanishing$f(x) = \max(0, x)$.
   b. Tanh (Hyperbolic Tangent): This function maps the input to the output in the range [−1,1][-1,1][−1,1] and aids in the learning of centralized data around zero. Formula:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

   c. Sigmoid: This function converts the input into an output in the range [0.1][0.1][1][0.1] and is often used for small-scale outputs. The formula:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

   Model Optimization[23]:

   a. Adam (Adaptive Moment Estimation): Combines the advantages of momentum and RMSProp, with parameter update formulas:

$$\theta_{\{t+1\}} = \theta_t - \frac{\eta}{\sqrt{\tilde{v}+s}} \tilde{m}_t \quad (5)$$

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

Where is the first moment estimated, and $\tilde{v}\tilde{m}_{tt}$ is the second moment estimated.

b. RMSprop: Mempertahankan learning rate stabil untuk parameter yang sering diperbarui, dengan rumus:

$$\theta_t = \theta_{\{t-1\}} = \theta_t - \frac{\eta}{\sqrt{E[g^2]t+s}} g_t \quad (6)$$

c. SGD (Stochastic Gradient Descent): Simple optimization method with parameter updates:

$$\theta_t = \theta_{\{t-1\}} - \eta \cdot \nabla_\theta J(\theta) \quad (7)$$

d. Adagrad, Adadelta, Nadam, and AdamW: Used to compare how different optimization algorithms affect the convergence and accuracy of the model.

4. Testing and Evaluation:

To evaluate the performance of prediction models, some commonly used evaluation metrics are *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), and *R-Squared* ($R2$). MAE measures the mean absolute difference between the predicted value and the actual value, which is formulated as [24]:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (8)$$

This metric provides an intuitive error measure without taking into account the direction of the error. In addition, RMSE is used to measure the mean of the squares of error between the predicted value and the actual value, which is formulated as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (9)$$

RMSE is more sensitive to large errors, thus giving a higher penalty on predictions that are far from the actual value. MAPE is used to measure the mean error in percentage form. MAPE is formulated as:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (10)$$

MAPE allows for the comparison of errors between different data scales, making it particularly useful in economic contexts. In addition, *the R-Squared* (R^2) metric is also used to measure how well the model can explain variations in actual data. R-Squared is formulated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (11)$$

where is the actual value, is the prediction value, and is the average value of all the actual values. Values range from 0 to 1. A value close to 1 indicates that the model can explain most variations in the data, while a value close to 0 indicates that the model cannot explain the variationsy_i $\hat{y}_i$ $\bar{y}R^2$.

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 905-922
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

3. RESULT

In this section, the research findings and hypothesis testing that had been done are presented. The research result data are elaborate and have visualizations like tables and graphs to create clear understanding. This part is a discussion addressing research questions, interpretation of findings and integration of research results into the existing knowledge in this section. Further, it can also be a part of discussions to create new theories or improve an old process based on the obtained results.

In addition, it contains the data pre-processing step which ensures that corruption is impossible on the data being used. The first step in this preprocess stage is missing value handling, using a linear interpolation method to complete the missing values. What I get after processing this missing value is present in table 2 (the process — on the data visualization in figure 3).

Tabel 2. Data Garlic Bulb

No	Date	Price
1	15/03/2021	35
2	16/03/2021	36.92
3	17/03/2021	35.25
4	18/03/2021	36.78
5	19/03/2021	36.78
6	20/03/2021	36.86
7	21/03/2021	37.36
8	22/03/2021	37.86
9	23/03/2021	37.19
10	24/03/2021	36.32
11	25/03/2021	38.42
12	26/03/2021	35
13	27/03/2021	36.91
14	15/03/2021	35
…	…	…
…	…	…
1333	06/11/2024	51.48
1334	07/11/2024	51.91

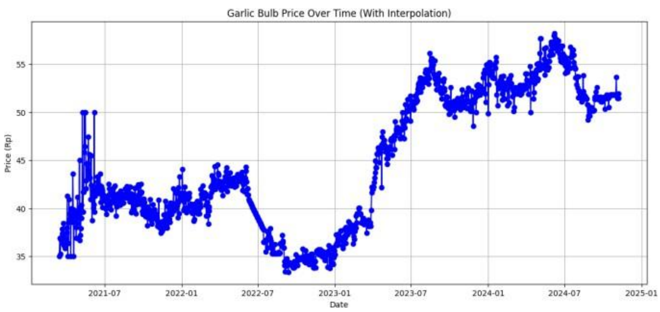

Figure 3. Mising Handling Results Graph

Figure 3 presents the results of handling missing values in the garlic bulb price dataset. The interpolated data ensures continuity in the time series, allowing the LSTM model to learn patterns more effectively without disruptions caused by missing entries.

Once we have dealt with missing values, we need to go ahead and normalize the data. We consider that it is necessary to perform this process in order to scale the data with a range of [0, 1], using the

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

MinMaxScaler method for scaling, intending at increasing the model stability during training and helping in convergence. Normalization is to ensure balance feature contribution to the model and a scale of larger features does not dominate. These normalized results make it easier for the LSTM model to identify patterns and trends within the data. Results of Data Normalization are shown in Figure 4.

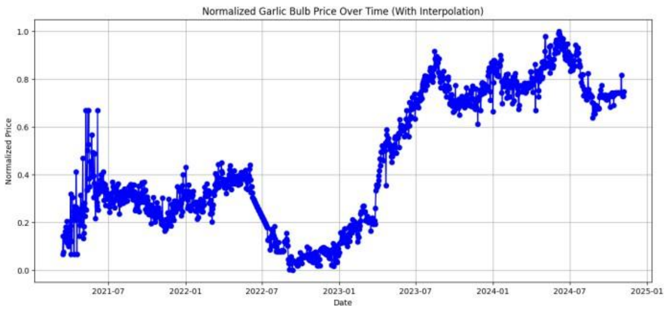

Figure 4. Graph of Data Normalization Results

Data transformation after normalization with MinMaxScaler method is shown in Figure 4. It transforms the data with a different range of values, into the range of [0, 1]. The significance of this normalization lies in the fact that uniform distributed data supports LSTM model by preventing numerical troubles while training (for example, convergence problems due to the size difference between features).

To weight features equally in model learning, normalization allows algorithm to learn better patterns from the given data. Moreover, it helps eliminate the bias that advanced features cause to model performance at a higher level. We can see from this visualization that the data structure and most of its information is preserved but it has been rescaled to prepare data with a dataset which could ensure stable and fast training of the models.

After normalization of the data, we can split the data into three parts; training, validation, and testing data. Training data, validation and test: 70–15–15 split. This separation is done to make sure that model has trained on sufficient data to learn the pattern, validated so we know about the performance during training and tested so we can evaluate model generalizations on never seen before data. This division of data and its resulting visuals are demonstrated in figure 5. This is important so that the model can get maximum performance and avoid overfitting, so that the prediction results are more accurate and reliable.

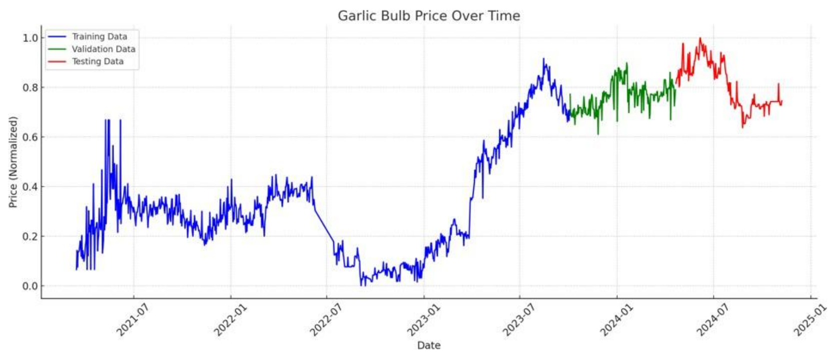

Figure 5. Splitting Data

The Garlic Bulb price data is first normalized, and the time series is then divided into training set, validation set and testing data.Set the division ratio accordingly to train different models Figure 5: Normalized Garlic Bulb price data used in this paper along with split of data sequence in three sets;

Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

training, validation or test of period over entire period under consideration. Training data is shown in blue, validation in green and testing in red. These visualizations can give us a good idea of the makeup of our data with respect to each data set ensuring that we have done good split between training and validation, testing on different data. This is important to reduce the chance of overfit that can prevent our model from producing reliable and accurate predictions on previously unseen data.

The following phase of this work, is to model the pre-treated data into a Long Short-Term Memory (LSTM) architecture. LSTM is an artificial recurrent neural network topology, ideally suited for time series with long-term dependencies. This particular model is well recognised to outperform the vanishing gradient problems of a traditional neural network such as RNN (Recurrent Neural Networks). LSTMs have the ability to retain relevant information for long time frames due to its cell state and gate mechanisms, which is useful since commodity prices often contain complex fluctuations and trends.

As the LSTM model is trained with 200 epochs in this step, it makes sure that the model gets sufficient time to learn all patterns and tendencies hidden within data. A large number of epochs has been selected to allow the model to potentially learn very long-term dependencies in the time series, which also have a significant impact on producing an accurate forecast. The model is trained using all possible combinations of activation functions and optimization algorithms, by testing each systematically to find the right fit.

Activation functions (ReLU, Sigmoid, Tanh; the one used in this work) ReLU (Rectified Linear Unit) can be quite efficient to break the gradient vanishing but it is generally only used for complex patterns recognizing data. Sigmoid, which is often designed for small scale output while Tanh makes it useful to obtain the data that goes around zero, hence making this a much balanced learning process as compared to when it makes use of the sigmoid function. By testing different activation functions, this study can better study the impact of function characteristics on LSTM predictive performance for commodity prices.

Other optimization algorithms are Adam, RMSprop, SGD, Adagrad, Adadelta and Nadam and AdamW. Adam and Nadam provides fast convergence, gradually adjusting to the different gradients while RMSprop offers each parameter a fixed learning rate depending on its update frequency. SGD is a straightforward optimization algorithm that provides steadiness; Adagrad and Adadelta feature adaptive parameter updates for incorporation. A total of 21 pairs were tested, from ReLU + Adam, Sigmoid + RMSprop to Tanh + AdamW did the combination look like for the LSTM model with an aim to identify the best configuration for predictions on price of Garlic Bulb.

Different ReLU activations with different optimizers At this stage, the combinations that are tested are: ReLU + Adam, ReLU + RMSprop, ReLU + SGD, ReLU + Adagrad, ReLU + Adadelta, Relu+Nadam and Relu+AdamW. ReLU was selected as the activation function because of computing efficiency and better avoidance of gradient vanishing problems. The set is paired in such a manner that it can study the performance of the model using speed or prediction accuracy. The results from this experiment appear in Figure 6 and provide a summary of the performance for each combination with respect to evaluation metric; Final Epoch Loss, RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error) & R^2 (determination coefficient). It provides a means of evaluating the accuracy and reliability of the model for all configurations tested.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

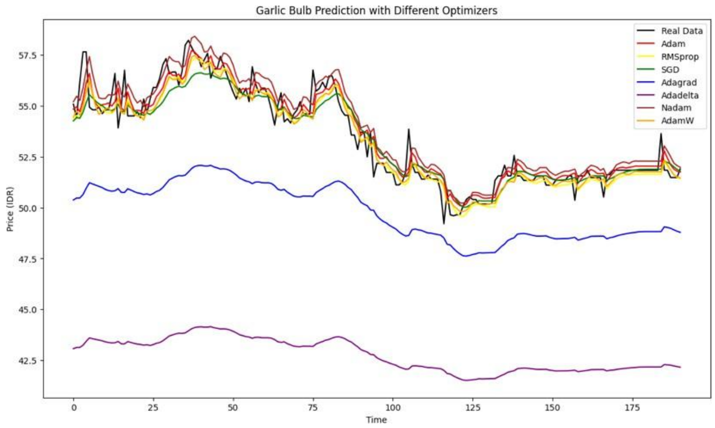

Figure 6. Relu Activation

Figure 6 shows the results of first experiment with ReLU as activation functions and seven different optimization algorithms which is also summarized in following results table. Figure 2: Status of the Garlic Bulb price prediction LSTM model generated by each combination (graph) and summary of its evaluation metric), Final Epoch Loss, RMSE, MAPE and R^2; giving a quick overview of how our subsequent configurations performed on the latter.

1. The best results were obtained from Adam, with a Final Epoch Loss of 0.001789, RMSE 0.701632, MAPE 0.009593 and $R^2 = 0.909794$ This indicates that ReLu + Adam is able to make accurate predictions with fairly low error and a fair amount of explainability for the high variation in data.
2. RMSprop gives to similar results as Adam where Final Epoch Loss 0.001949 and RMSE 0.723185, MAPE 0.009454 and R^2 is about.904167 which means it again the model is pretty reliable in catching the data pattern here too.
3. SGD: Final Epoch Loss= 0.001975 RMSE = 0.797999, MAPE= 0.010413 $R^2 = 0.883314$ While more stable, SGD showed lower accuracy than Adam and RMSprop.
4. However, it shows worse results than Adagrad with a higher Final Epoch Loss(0.005975), and RMSE (3.828303), MAPE(0.067511) and negative R^2(-1.685518); which indicates poor and suboptimal performance for this prediction.
5. Adadelta performed the worst: with Final Epoch Loss 0.034092, RMSE 10.814232, MAPE 0.199155 and R^2 -20.429242 over all epochs The implication of these results is that this algorithm does not perform well enough to use in this combination at all.
6. Final Epoch Loss 0.001801, RMSE 0.851298, MAPE 0.012926, R^2 of 0.867206 Now we see that Nadam performed relatively well, albeit not well above Adam and RMSprop the model has some ability to make accurate predictions on this dataset
7. Final Epoch Loss 0.001791, RMSE 0.712480, MAPE 0.009079 and R^2 0.906983 (AdamW actually performed pretty much as Adam) Thus, confirms that AdamW is solid choice where results are on par to Adam.

The output of this experiment : Best among all: ReLU + Adam (Final Epoch Loss = 0.001789, RMSE 0.701632, MAPE = 0.009593 and $R^2 = 0.909794$). This indicates that this combination can make very accurate predictions with low error and a great deal of explained variation in the data. Therefore, It can be concluded that ReLU + Adam is the most optimal configuration for model LSTM to predict price of Garlic Bulb commodity in this experiment with a fast convergence speed and decent prediction result on time series data.

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

In the experiment below, to evaluate how the model performed on total Twitter data with Sigmoid activation function together with different optimization algorithms. All tested combinations include Sigmoid + Adam, Sigmoid + RMSprop, Sigmoid + SGD, Sigmoid + Adagrad, Signoid+ Adadelta, Sigmoide+Nadam and Signoid_AdamW. The Sigmoid activation function was selected due to its properties of mapping input to output in the range [0, 1], which are useful for small data and ensures output is kept within an interpretable bound. The evaluation time then showed the most important and prominent results, for example in this experiment was Sigmoid + Adam combination with Final Epoch Loss, RMSE, MAPE, and R^2 at values that are lower than other combination processes this indicates that it has a fairly good level of accuracy. Though the Sigmoid + RMSprop & Sigmoid + AdamW also represent competitive combinations, they only outperform Sigmoid + Adam by margin of 1% in accuracy. In contrast, Sigmoid + Adagrad and Sigmoid + Adadelta experience performance failures once again with a less than R^2 score of 0.2 and higher RMSE which illustrates that this algorithm is non-optimal for the model containing the Sigmoid activation function in LSTM. Figure 7 demonstrates the anecdotal results of this experiment, showing a comparison of performance for each combination in predicting commodity prices.

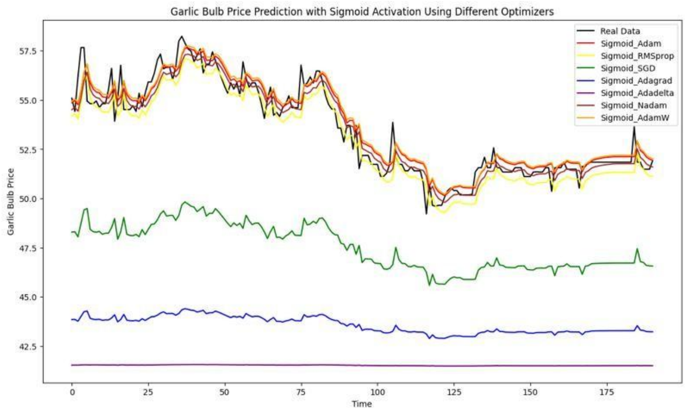

Figure 7. Sigmoid Activation

The second experiment uses the Sigmoid activation function with various optimization algorithms shown in Figure 7 and summarized in the following results table. The graph shows the Garlic Bulb price prediction generated by the LSTM model with this combination, while the table summarizes the evaluation metrics, namely Final Epoch Loss, RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and R¬≤ (determination coefficient), which shows the model's performance.

1. Sigmoid + Nadam showed the best results among the Sigmoid activation function combinations, with a Final Epoch Loss of 0.002051, RMSE 0.705506, MAPE 0.009186, and R¬≤ of 0.908795. These results show the model's high accuracy and ability to explain large variations in data.
2. Sigmoid + Adam has a similar performance to Nadam, with a Final Epoch Loss of 0.002070, RMSE of 0.738015, MAPE of 0.010514, and an R¬≤ of 0.900197, indicating that this combination remains one of the solid options.
3. Sigmoid + RMSprop yielded a Final Epoch Loss of 0.002500 and RMSE of 0.849391, with a MAPE of 0.012214 and an R¬≤ of 0.867800. Although his performance is lower than Nadam and Adam, this combination still shows quite good results.
4. Sigmoid + AdamW recorded a Final Epoch Loss of 0.002093, RMSE 0.780038, MAPE 0.011504, and R¬≤ 0.888507, indicating a fairly good performance, although not as good as Nadam.
5. Sigmoid + SGD, Adagrad, and Adadelta performed well below expectations. Sigmoid + SGD has a Final Epoch Loss of 0.012433 and an RMSE of 6.047754, with a negative R¬≤ of -5.701988, indicating a very high error. Adagrad and Adadelta recorded Final Epoch Loss of 0.030631 and

Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

0.044292, respectively, with R¬≤ of -17.549844 and -26.054401, indicating that this combination is not optimal for use in LSTM models with Sigmoid activation functions.

The results of the second experiment showed that Sigmoid + Nadam was the best. This combination achieved a Final Epoch Loss of 0.002051, RMSE 0.705506, MAPE 0.009186 and an R^2 of 0.908795. The LSTM model which was using Sigmoid activation function and Nadam optimization algorithm is capable of offering with high accuracy predictions with less error as well as a good data variations explanation score. Among other configurations with Sigmoid activation function, this combination indicates faster convergence and high reliability which makes it the best choice for time series data analysis in this experiment.

The third experiment was done using LSTM model with Tanh activation function and different optimization algorithms to compare the performance of the model. Introduced as a more powerful alternative to the Sigmoid activation, Tanh gives output in [-1, 1] which means that data is naturally centred around zero so continuous time series data can be learned more effectively by the model. In this stage we tested the combination of these optimization algorithms: Tanh + Adam, Tanh + RMSprop, Tanh + SGD, Tanh + Adagrad, Tanh + Adadelta, Tanh + Nadam and finally Tanh + AdamW. We then compute Final Epoch Loss, RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and R^2(superficial meaning of the r^2) by evaluating each of these combinations. As the results of this trial can be seen in Figure 8 which shows a comparison between using each combination with the performance result on the price prediction of The Garlic Bulb commodity.

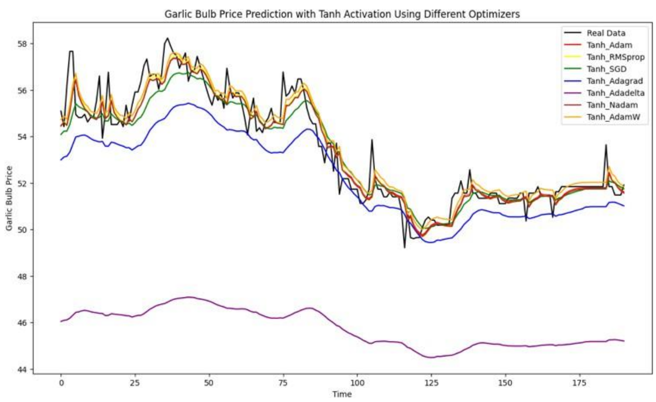

Figure 8. Tanh Activation

The results of the third experiment using the Tanh activation function with various optimization algorithms are shown in Figure 8 and summarized in the following results table. The graph shows the Garlic Bulb price prediction generated by the LSTM model with this combination, while the table summarizes the evaluation metrics, namely Final Epoch Loss, RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and R¬≤ (determination coefficient), which shows the model's performance.

1. Tanh + Adam showed the best results with a Final Epoch Loss of 0.002025, RMSE 0.705124, MAPE 0.009174, and R¬≤ of 0.908894. These results show that this combination is able to provide highly accurate predictions with low error and the ability to explain high data variations.
2. Tanh + RMSprop has almost equivalent performance to Tanh + Adam, with a Final Epoch Loss of 0.001989, RMSE 0.712449, MAPE 0.009276, and R¬≤ of 0.906992. This combination demonstrates reliable prediction ability.
3. Tanh + Nadam also performed very well with Final Epoch Loss 0.001966, RMSE 0.708685, MAPE 0.009246, and R¬≤ 0.907971, indicating that the algorithm is a solid choice.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

4. Tanh + AdamW recorded a Final Epoch Loss of 0.002000, RMSE 0.718854, MAPE 0.009894, and R¬≤ 0.905312, showing a fairly good performance although not as good as Tanh + Adam.
5. Tanh + SGD showed lower results with Final Epoch Loss 0.002051, RMSE 0.824817, MAPE 0.010781, and R¬≤ 0.875339, signaling lower accuracy compared to the best combinations.
6. Tanh + Adagrad and Tanh + Adadelta performed well below expectations, with a high RMSE and a negative R¬≤. Adagrad recorded a Final Epoch Loss of 0.002477, RMSE 1.447369, and R¬≤ of 0.616139, while Adadelta had a Final Epoch Loss of 0.016653, RMSE 7.864768, and R¬≤ -10.334104, indicating the model's inability to predict data with sufficient accuracy.

Among these results, Tanh + Adam is the best among this experiment followed by Tanh + RMSprop and Tanh + Nadam that have relatively fast convergence rate also accurate prediction result. Table 3 will present all of the results from experiments using other combinations of activation functions and optimization algorithms, namely: ReLU, Sigmoid and Tanh activation function and Adam, RMSprop, Nadam as well as others. In this table we can have an overview of important evaluation metrics (Final Epoch Loss, RMSE, MAPE and R^2 (determination coefficient)) to know how well each combination performed. Comparative analysis can be done more effectively through this presentation, which helps to devise the best configuration of LSTM model in predicting the price of commodity Garlic Bulb.

Table 3. Model Comparison Results

No	Activation	Optimizer	Final Epoch Loss	RMSE	MAPE	R2
1	Relu	Adam	0.001789	0.701632	0.009593	0.909794
2		RMSprop	0.001949	0.723185	0.009454	0.904167
3		SGD	0.001975	0.797999	0.010413	0.883314
4		Adagrad	0.005975	3.828303	0.067511	-1.685518
5		Adadelta	0.034092	10.814232	0.199155	-20.429242
6		Nadam	0.001801	0.851298	0.012926	0.867206
7		AdamW	0.001791	0.71248	0.009079	0.906983
8	Sigmoid	Adam	0.00207	0.738015	0.010514	0.900197
9		RMSprop	0.0025	0.849391	0.012214	0.8678
10		SGD	0.012433	6.047754	0.109813	-5.701988
11		Adagrad	0.030631	10.061493	0.183475	-17.549844
12		Adadelta	0.044292	12.150978	0.221704	-26.054401
13		Nadam	0.002051	0.705506	0.009186	0.908795
14		AdamW	0.002093	0.780038	0.011504	0.888507
15	Tanh	Adam	0.002025	0.705124	0.009174	0.908894
16		RMSprop	0.001989	0.712449	0.009276	0.906992
17		SGD	0.002051	0.824817	0.010781	0.875339
18		Adagrad	0.002477	1.447369	0.021725	0.616139
19		Adadelta	0.016653	7.864768	0.142905	-10.334104
20		Nadam	0.001966	0.708685	0.009246	0.907971
21		AdamW	0.002	0.718854	0.009894	0.905312

Table 3 summarizes the results of three sets of LSTM model experiments with various combinations of activation functions and optimization algorithms, evaluated using Final Epoch Loss, RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and R^2 (determination coefficient). Each combination is assessed to determine the most effective configuration for predicting garlic bulb prices. Among the tested combinations, ReLU + Adam achieves the best performance, with Final Epoch Loss of 0.001789, RMSE of 0.701632, MAPE of 0.009593, and R^2 of 0.909794. This superior performance can be attributed to ReLU's efficiency in avoiding vanishing gradient problems

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

and Adam's adaptive learning rate and momentum-based updates, which contribute to stable and fast convergence. ReLU is particularly effective in deep networks because it does not saturate for positive inputs, allowing gradients to propagate efficiently. Meanwhile, Adam combines the benefits of momentum (like RMSprop) and adaptive learning rate scaling (like Adagrad), making it well-suited for optimizing deep learning models with non-stationary objectives, such as time-series forecasting. The combination of these two allows the model to learn complex patterns more effectively, leading to better generalization on unseen data.

On the other hand, other optimizer and activation function combinations show limitations. For instance, ReLU + Adadelta performs the worst, with RMSE of 10.814232 and R^2 of -20.429242, suggesting that Adadelta struggles with stability and weight updates in this context. Unlike Adam, Adadelta does not maintain a momentum term, which can lead to slower convergence and poor generalization, especially when dealing with highly dynamic price fluctuations. For the Sigmoid activation function, the best combination is Sigmoid + Nadam, with RMSE of 0.705506 and R^2 of 0.908795. However, Sigmoid suffers from saturation issues, where large positive or negative inputs push gradients close to zero, potentially slowing down learning. While Nadam (a momentum-accelerated variant of Adam) helps mitigate this by improving gradient updates, the Sigmoid activation function still limits overall performance compared to ReLU.

The Sigmoid + Adadelta combination performs the worst, with RMSE of 12.150978 and R^2 of -26.054401, reinforcing the instability of Adadelta in this setting. For the Tanh activation function, the best combination is Tanh + Adam, with RMSE of 0.705124 and R^2 of 0.908894. Tanh offers an advantage over Sigmoid by centering outputs around zero, which helps with gradient propagation in deep networks. However, compared to ReLU, Tanh still faces saturation issues, making it slightly less effective in deep learning architectures. The worst-performing combination in this group is Tanh + Adadelta, with RMSE of 7.864768 and R^2 of -10.334104, further confirming Adadelta's inefficiency in optimizing LSTM models. In summary, ReLU + Adam outperforms other combinations due to ReLU's ability to efficiently propagate gradients and Adam's stability in weight updates, leading to faster convergence and better generalization. In contrast, Adadelta consistently underperforms across all activation functions, indicating its limitations in handling dynamic time-series data. These insights suggest that selecting the right activation function and optimizer is crucial for maximizing LSTM performance in commodity price prediction.

From the three sets of experiments, the best performance results of each activation function will be displayed in a graphical visualization in Figure 9 to facilitate visual comparison and analysis.

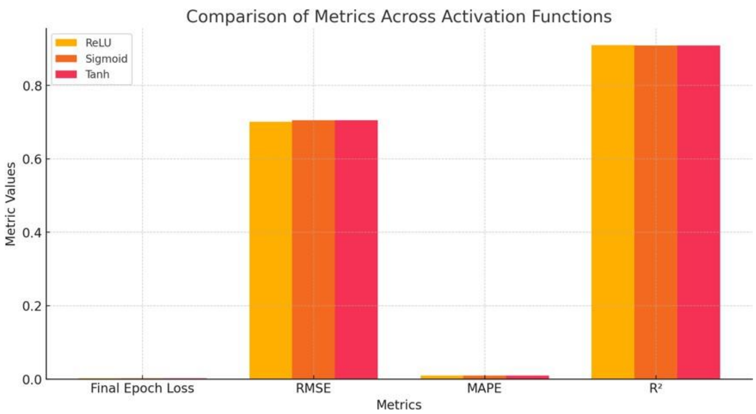

Figure 9. Best Results Every Activation

 Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

In Fig. 9, the best combination of these three activation functions is ReLU, Sigmoid and Tanh for Final Epoch Loss, RMSE, MAPE and R^2 performance metrics in LSTM model compared with each other. Results showed that ReLU + Adam combination with the highest overall performance between other comparison, accomplished a Final Epoch Loss of 0.001789, RMSE 0.701632, MAPE 0.009593 and $R^2 = 0.909794$ which indicates high accuracy along with explaining almost 91% of data variability by this model respectively. Similar, but slightly worse than ReLU + Adam, was the performance of Sigmoid + Nadam – 2nd place with Final Epoch Loss 0.002051 | RMSE 0.705506 | MAPE 0.009186 | R^2 0.908795 Like the Sigmoid + Nadam combination, the Tanh + Adam combination yielded a Final Epoch Loss of 0.002025, RMSE 0.705124, MAPE 0.009174, and R^2 0.908894 which are fairly similar results but with an advantage in MAPE over the previous one. In conclusion, the combination of using ReLU + Adam is the best results we achieved to predict the price of Garlic Bulb. commodity, as it has the lowest RMSE value, the highest R^2, and the lowest Final Epoch Loss among all the combinations teste.

4. DISCUSSIONS

This study aims to identify the optimal combination of activation functions and optimization algorithms that can improve the accuracy of the Long Short-Term Memory (LSTM) model in predicting garlic commodity prices. The three activation functions tested were ReLU, Sigmoid, and Tanh, which were paired with various optimization algorithms such as Adam, RMSprop, SGD, Nadam, Adadelta, and AdamW. The results showed that the ReLU + Adam combination gave the best results with an RMSE of 0.701632 and an R^2 of 0.909794, indicating that this combination was able to generate accurate predictions and explain most of the data variations. ReLU is effective in dealing with disappearing gradient issues, while Adam provides stability and fast convergence. This result is in line with previous research which confirms that Adam is an efficient and stable optimization algorithm to improve LSTM performance in complex time series data prediction [1], [2].

In addition, the Sigmoid + Nadam combination shows excellent performance with an RMSE of 0.705506 and an R^2 of 0.908795. Nadam, which combines the properties of momentum-based and adaptive optimization, has been shown to improve the speed of convergence and prediction accuracy, in line with the results of previous studies [1]. The combination of Tanh + Adam also provides high performance with an RMSE of 0.705124 and an R^2 of 0.908894, confirming that Tanh is suitable for use on data centered around zero and is capable of capturing complex time series data patterns [3], [7]. Previous studies have also shown that the use of LSTM with Tanh is more effective than other prediction methods such as ARIMA in dealing with nonlinear data fluctuations [6], [7].

These results show that the combination of ReLU + Adam is the best configuration for prediction of nonlinear and fluctuating data, but the combination of Sigmoid + Nadam and Tanh + Adam is also relevant to use in certain situations. This supports research that mentions the importance of selecting momentum-based optimization algorithms to improve the stability of prediction models [1], [2]. For further development, future research may consider external factors such as changes in government policies and global market conditions to improve the accuracy of predictions [9], [25]. In addition, the incorporation of hybrid methods such as genetic algorithm-based regression can provide more accurate and stable prediction results, as has been proven by previous studies [5], [11].

5. CONCLUSION

This study demonstrates that the selection of an optimal activation function and optimization algorithm significantly impacts the performance of LSTM models in predicting garlic bulb commodity prices. Various combinations were tested, and results indicate that ReLU + Adam provides the best performance, achieving the lowest Final Epoch Loss (0.001789), RMSE (0.701632), MAPE (0.009593),

Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

and the highest R^2 (0.909794). These findings align with previous research that identifies Adam as a reliable and efficient optimization technique. Other high-performing configurations include Sigmoid + Nadam and Tanh + Adam, highlighting the benefits of momentum-based optimization algorithms in improving LSTM model accuracy. The strong performance of Tanh + Adam also suggests that Tanh is advantageous when dealing with data centered around zero. Certain activation function and optimizer combinations offer optimal performance for time series data analysis, broadening the predictive modeling approaches available for commodity price forecasting.

The results of this study provide valuable insights for stakeholders in the agricultural sector, policymakers, and market analysts in developing more accurate price prediction systems. By leveraging the best-performing LSTM configurations identified, businesses and governments can enhance decision-making processes, optimize supply chain strategies, and mitigate risks associated with price volatility. Further research can explore hybrid LSTM architectures, integrating external economic indicators, policy changes, weather conditions, and global market trends to improve price prediction accuracy. Additionally, experimenting with deeper LSTM models with multiple hidden layers or combining LSTM with other deep learning architectures such as GRUs (Gated Recurrent Units) may further enhance predictive performance. Testing the model on other agricultural commodities is also recommended to assess its generalizability across different market conditions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENT

The researcher expressed his gratitude to the Institute for Research and Community Service (LPPM) Telkom University Purwokerto for the funding support and facilities that have been provided during the implementation of this research. This assistance and support is an important part of the smooth research process until the completion of this report. Hopefully the results of this research can make a positive contribution to the development of science and its application in society.

REFERENCES

[1] Y. Setiawan, T. Tarno, and P. Kartikasari, ‘PREDIKSI HARGA JUAL KAKAO DENGAN METODE LONG SHORT-TERM MEMORY MENGGUNAKAN METODE OPTIMASI ROOT MEAN SQUARE PROPAGATION DAN ADAPTIVE MOMENT ESTIMATION DILENGKAPI GUI RSHINY’, *J. Gaussian*, vol. 11, no. 1, pp. 99–107, May 2022, doi: 10.14710/j.gauss.v11i1.33994.

[2] J. Wang and J. Wang, ‘A New Hybrid Forecasting Model Based on SW-LSTM and Wavelet Packet Decomposition: A Case Study of Oil Futures Prices’, *Comput. Intell. Neurosci.*, vol. 2021, no. 1, p. 7653091, Jan. 2021, doi: 10.1155/2021/7653091.

[3] X. W. GAO and A. GAO, ‘COVID-CBR: A Deep Learning Architecture Featuring Case-Based Reasoning for Classification of COVID-19 from Chest X-Ray Images’, *2021 20th IEEE Int. Conf. Mach. Learn. Appl. ICMLA*, pp. 1319–1324, 2022, doi: 10.1109/ICMLA52953.2021.00214.

[4] E. J. Griffis, O. S. Patil, Z. I. Bell, and W. E. Dixon, ‘Lyapunov-Based Long Short-Term Memory (Lb-LSTM) Neural Network-Based Control’, *IEEE Control Syst. Lett.*, vol. 7, pp. 2976–2981, 2023, doi: 10.1109/LCSYS.2023.3291328.

[5] N. Pai, I. Velchamy, and N. B. Ramesh, ‘Forecasting of Multistep Multivariate Financial Data Through GSO Algorithm Infused Vanilla LSTM Model’, *J. Comput. Sci.*, vol. 19, no. 7, pp. 909–924, Jul. 2023, doi: 10.3844/jcssp.2023.909.924.

Vol. 6, No. 2, April 2025, Page. 905-922
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4412

[6] I. sibel Kervanci and F. Akay, 'LSTM Hyperparameters optimization with Hparam parameters for Bitcoin Price Prediction', *Sak. Univ. J. Comput. Inf. Sci.*, vol. 6, no. 1, pp. 1–9, Apr. 2023, doi: 10.35377/saucis...1172027.

[7] Y. Loday, P. Apirukvorapinit, and P. Vejjanugraha, 'Stock Price Prediction Using Modified Bidirectional Long Short-Term Memory and Deep Learning Models: A Case Study of Bhutan Tourism Corporation Limited Stock Data', in *2023 8th International Conference on Business and Industrial Research (ICBIR)*, Bangkok, Thailand: IEEE, May 2023, pp. 645–650. doi: 10.1109/ICBIR57571.2023.10147486.

[8] N. Chen, 'Exploring the development and application of LSTM variants', *Appl. Comput. Eng.*, vol. 53, no. 1, pp. 103–107, Mar. 2024, doi: 10.54254/2755-2721/53/20241288.

[9] P. R. Low and E. Sakk, 'Comparison between autoregressive integrated moving average and long short term memory models for stock price prediction', *IAES Int. J. Artif. Intell. IJ-AI*, vol. 12, no. 4, p. 1828, Dec. 2023, doi: 10.11591/ijai.v12.i4.pp1828-1835.

[10] Q. Liang, Y.-C. Chan, P. B. Changala, D. J. Nesbitt, J. Ye, and J. Toscano, 'Ultrasensitive multispecies spectroscopic breath analysis for real-time health monitoring and diagnostics', *Proc. Natl. Acad. Sci.*, vol. 118, no. 40, p. e2105063118, Oct. 2021, doi: 10.1073/pnas.2105063118.

[11] S.-F. Yang, S.-W. Choi, and E.-B. Lee, 'A Prediction Model for Spot LNG Prices Based on Machine Learning Algorithms to Reduce Fluctuation Risks in Purchasing Prices', *Energies*, vol. 16, no. 11, p. 4271, May 2023, doi: 10.3390/en16114271.

[12] Y. Lu, Y. Teng, Q. Zhang, and J. Dai, 'Prediction Model for the Chemical Futures Price Using Improved Genetic Algorithm Based Long Short-Term Memory', *Processes*, vol. 11, no. 1, p. 238, Jan. 2023, doi: 10.3390/pr11010238.

[13] G. Huang, 'Missing data filling method based on linear interpolation and lightgbm', *J. Phys. Conf. Ser.*, vol. 1754, no. 1, p. 012187, Feb. 2021, doi: 10.1088/1742-6596/1754/1/012187.

[14] X. Zhang, C. Li, K.-L. Chen, D. Chrysostomou, and H. Yang, 'Stock Prediction with Stacked-LSTM Neural Networks', in *2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, Hainan, China: IEEE, Dec. 2021, pp. 1119–1125. doi: 10.1109/QRS-C55045.2021.00166.

[15] O. S. Alshehri, O. M. Alshehri, and H. Samma, 'Blood Glucose Prediction Using RNN, LSTM, and GRU: A Comparative Study', in *2024 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, Hammamet, Tunisia: IEEE, Apr. 2024, pp. 1–5. doi: 10.1109/IC_ASET61847.2024.10596176.

[16] T. J. Mbah, H. Ye, J. Zhang, and M. Long, 'Using LSTM and ARIMA to Simulate and Predict Limestone Price Variations', *Min. Metall. Explor.*, vol. 38, no. 2, pp. 913–926, Apr. 2021, doi: 10.1007/s42461-020-00362-y.

[17] Y. Chen, 'Price Prediction of Cude oil, Gold, and Cotton Based on OLS, Random Forest, and Lightgbm', *BCP Bus. Manag.*, vol. 38, pp. 3361–3369, Mar. 2023, doi: 10.54691/bcpbm.v38i.4300.

[18] F. Tang, H. Yi, H. Cui, and K. Zhang, 'A Stock Prediction Method Based on Singular Spectrum Analysis and Long Short-Term Memory Network', in *2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT)*, Qingdao, China: IEEE, Jul. 2023, pp. 967–970. doi: 10.1109/ICEICT57916.2023.10245558.

[19] J. Lv, C. Wang, W. Gao, and Q. Zhao, 'An Economic Forecasting Method Based on the LightGBM-Optimized LSTM and Time-Series Model', *Comput. Intell. Neurosci.*, vol. 2021, no. 1, p. 8128879, Jan. 2021, doi: 10.1155/2021/8128879.

[20] V. G. Kowti, 'Stock Price Prediction using LSTM and KNN Algorithms', *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 9, pp. 1591–1595, Sep. 2023, doi: 10.22214/ijraset.2023.55894.

[21] S. Sinsomboonthong, 'Performance Comparison of New Adjusted Min-Max with Decimal Scaling and Statistical Column Normalization Methods for Artificial Neural Network Classification', *Int. J. Math. Math. Sci.*, vol. 2022, pp. 1–9, Apr. 2022, doi: 10.1155/2022/3584406.

[22] M. H. Essai Ali, A. B. Abdel-Raman, and E. A. Badry, ‘Developing Novel Activation Functions Based Deep Learning LSTM for Classification’, *IEEE Access*, vol. 10, pp. 97259–97275, 2022, doi: 10.1109/ACCESS.2022.3205774.

[23] M. Taqiyuddin, K. Adi, O. Dwi Nurhayati, and H. Ochi, ‘Comparison of Optimizers for Drone Signal Detection Using Convolutional Neural Networks (CNN)’, *E3S Web Conf.*, vol. 448, p. 02025, 2023, doi: 10.1051/e3sconf/202344802025.

[24] D. Chicco, M. J. Warrens, and G. Jurman, ‘The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation’, *PeerJ Comput. Sci.*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[25] H. Xu, J. Liang, and W. Zang, ‘Prediction of Agricultural Commodities Futures Prices: A DQN-LSTM Method’, Nov. 30, 2021. doi: 10.21203/rs.3.rs-1097759/v1.