BAREKENG: Journal of Mathematics and Its Applications
March 2025 Volume 19 Issue 1 Page 0245–0258
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss1pp0245-0258

THE COMPARISON OF LONG SHORT-TERM MEMORY AND BIDIRECTIONAL LONG SHORT-TERM MEMORY FOR FORECASTING COAL PRICE

Indra Rivaldi Siregar[1*], Adhiyatma Nugraha[2], Khairil Anwar Notodiputro[3], Yenni Angraini[4], Laily Nissa Atul Mualifah[5]

[1,2,3,4,5]*Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University*
Jl Meranti, Bogor, 16680, Indonesia

*Corresponding author's e-mail: * **siregarindra@apps.ipb.ac.id***

ABSTRACT

Article History:

Received: 28th May 2024
Revised: 31st August 2024
Accepted: 2nd September 2024
Published: 13th January 2025

Keywords:

BiLSTM;
Coal Price;
LSTM;
Time Series;
Walk Forward Validation.

Coal remains vital for global energy despite recent demand fluctuations due to the COVID-19 pandemic and geopolitical tensions. The International Energy Agency (IEA) projected a decline in global coal demand starting in early 2024, driven by increasing renewable energy adoption. As one of the top coal exporters, Indonesia must adjust to these changes. This study aims to forecast future coal prices using historical data from Indonesia's Ministry of Energy and Mineral Resources (KESDM), applying and comparing Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) models. While BiLSTM has shown advantages in other contexts and studies, its effectiveness for coal price forecasting remains underexplored. To ensure robust predictions, we employ walk-forward validation, which divides the data into six segments and evaluates 90 hyperparameter combinations across all segments. The BiLSTM model consistently outperforms the LSTM model, achieving lower average RMSE and MAPE values. Specifically, BiLSTM records an average MAPE of 7.847 and RMSE of 10.485, compared to LSTM's 10.442 and 11.993, respectively. The Diebold-Mariano (DM) test using squared error and absolute error loss functions further corroborates these findings, with most segments showing significant improvements in favor of BiLSTM, indicated by negative DM-test statistics and p-values below 0.01 or 0.10. This superior performance continues into the testing data, where BiLSTM maintains lower error metrics and a significant result of the DM test, underscoring its reliability for forecasting. In the final stage, the forecasts from both models indicate a nearly linear downward trend in coal prices over the next 18 months, aligning with the International Energy Agency's 2023 projection of a structural decline in coal demand driven by the sustained growth of clean energy technologies.

How to cite this article:

I. R. Siregar, A. Nugraha, K. A. Notodiputro, Y. Angraini and L. N. A. Mualifah., "THE COMPARISON OF LONG SHORT-TERM MEMORY AND BIDIRECTIONAL LONG SHORT-TERM MEMORY FOR FORECASTING COAL PRICE," *BAREKENG: J. Math. & App.*, vol. 19, iss. 1, pp. 0245-0258, March, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Coal plays a significant role in the global economy as it is the primary energy source for power generation, steelmaking, cement production, and other industrial sectors. Despite the turmoil of recent years, such as the drastic drop in demand during the COVID-19 pandemic and the sharp spike post-recovery from the pandemic, the global coal market still has a strong impact. The conflict between Russia and Ukraine has prompted European countries to rely on coal as an energy alternative due to soaring natural gas prices [1]. Despite controversial due to its environmental impact and market uncertainty, coal remains an integral part of today's energy infrastructure.

In 2023, according to the International Energy Agency (IEA), they projected a decline in global demand for coal beginning in early 2024. One of the factors that could influence the decline in coal consumption globally is the increasing adoption of renewable energy compared to the overall growth in electricity demand [2][3]. Indonesia, one of the largest coal exporters today and even the world's largest producer in 2022, must plan strategic steps and policies accordingly in light of this changing trend. According to information from the Ministry of Energy and Mineral Resources (KESDM), the reference coal prices for the last three months since December 2023 have declined in the range of 117.38 - 124.95 USD/ton [4]. Therefore, this study aims to utilize historical data of reference coal prices published by KESDM to develop a time series model that can provide accurate forecasting for the future.

Various methods have been applied, including deep learning techniques such as Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) for time series forecasting. LSTM is a development of a Recurrent Neural Network (RNN) equipped with specialized memory cells to store information in the long term. This technology helps overcome the vanishing gradient problem often occurring in RNNs when processing long sequential data [5]. Gao et al. (2018) found that LSTM outperforms moving average (MA), exponential moving average (EMA), and support vector machine (SVM) methods in stock closing price prediction, achieving MAPE values below 0.73 for 100, 200, and 400-day scenarios [6]. Additionally, several other studies have similarly demonstrated LSTM's superiority in various case studies compared to alternative methods [7][8][9]. On the other hand, BiLSTM is an advanced development of LSTM that consists of two layers, a forward layer to process previous information and a backward layer to process subsequent information [10]. BiLSTM enables more intricate and precise sequential data modeling as it can simultaneously access context from the past and the future [11].

Based on previous works, it is reasonable to assert that BiLSTM is generally superior to LSTM. Siami-Namini (2019) found that BiLSTM outperforms LSTM, achieving a 37.78% reduction in error rate for stock price prediction [12]. Other studies support this conclusion, showing BiLSTM's superiority over LSTM [10] [13]. However, despite BiLSTM's advantages in these contexts, it is crucial to compare both models to understand their performance in different scenarios, particularly for coal price forecasting. For instance, Handhayani (2022) demonstrated no significant difference between LSTM and BiLSTM in forecasting the air pollution index in Jakarta [14].

In the coal energy sector, many time series models also have been performed for many different purposes, such as forecasting coal prices for the future. For example, ARIMA and Transfer Functions are utilized for forecasting coal prices, with the Transfer Functions as the best model by comparing MAPE values [15]. Another study also combined the ARIMA and SVM models to forecast coal prices in the Chinese market to reduce the error generated by a single model. The combined results of the two models show a MAPE value of 1.49%, lower than the single ARIMA and SVM models [16]. Liu (2021) used three main categories, namely logistics factors, demand and supply impacts, and market factors, to forecast coal prices and applied an LSTM neural network model with an enhanced Adam optimizer, resulting in higher accuracy with an MAE of approximately 0.5 [17]. Xu & Zhang (2022) also applied the non-linear auto-regressive neural networks in forecasting the daily closing price of thermal coal traded on the China Zhengzhou Commodity Exchange. The study performed well, with relative RMSE values of 1.48, 1.49, and 1.47 for training, validation, and testing data, respectively [18]. Several other studies have employed deep learning for coal price forecasting [19] [20]. However, to the authors' knowledge, no research has yet compared BiLSTM to LSTM for coal price forecasting.

Therefore, this study compares LSTM and BiLSTM models in forecasting coal prices. The challenge of both methods is adjusting hyperparameters to produce the optimal models. We use a walk-forward validation scenario to get the best hyperparameter for LSTM and BiLSTM models. It is also helpful in increasing our confidence about the steady prediction in our constructed models. We also apply the Diebold-Mariano test [21] to compare both models to provide relevant claims about the best performance model.

2. RESEARCH METHOD

2.1 Research Data

Based on Figure 1, the data in this study consists of reference coal prices (in USD per ton) from the Ministry of Energy and Mineral Resources (ESDM) obtained from the website minerba.esdm.go.id/harga_acuan. The data comprised monthly period data from January 2009 to January 2024, totaling 181 observations [4].

Figure 1. Coal Price Time Series Data
Source: *Ministry of Energy and Mineral Resources Indonesia*

2.2 Steps of Data Analysis

Univariate time series modelling is the focus of this work. Initially, the data is transformed into a supervised learning framework, where y_{t-1} serves as the predictor. The subsequent step involves designating the last 1.5 years of data as the testing set and the remaining data as the training set. After that, the entire dataset is normalized by applying Equation (1). The purpose of min-max normalization is to scale the data within a range of [0, 1], ensuring that all features contribute equally to the model and improving the model performance [22] [23] .

$$y_{norm} = \frac{y - y_{min}}{y_{max} - y_{min}} \tag{1}$$

In the walk-forward validation scenario, the training set is further divided into training data (for model building) and validation data (for model evaluation), with a sliding window width set at one year. The hyperparameter tuning is performed on each piece (segment) of data in the walk-forward validation scenario, and these hyperparameters can determine the level of reliability and performance in the model [24]. We expect that this approach makes robust models as they are assessed across different validation data sets. After obtaining the best hyperparameters, the model is rebuilt using the entire training set (including validation data), and then we evaluate it on the testing set. The testing set plays a crucial role as it represents real-world data assumed to be previously unknown. In the final stage, the entire dataset (training set and testing set) is utilized to construct the final model for forecasting coal prices over the next 1.5 years (18 months). The detailed steps of the procedure in this study can be seen in Figure 2.

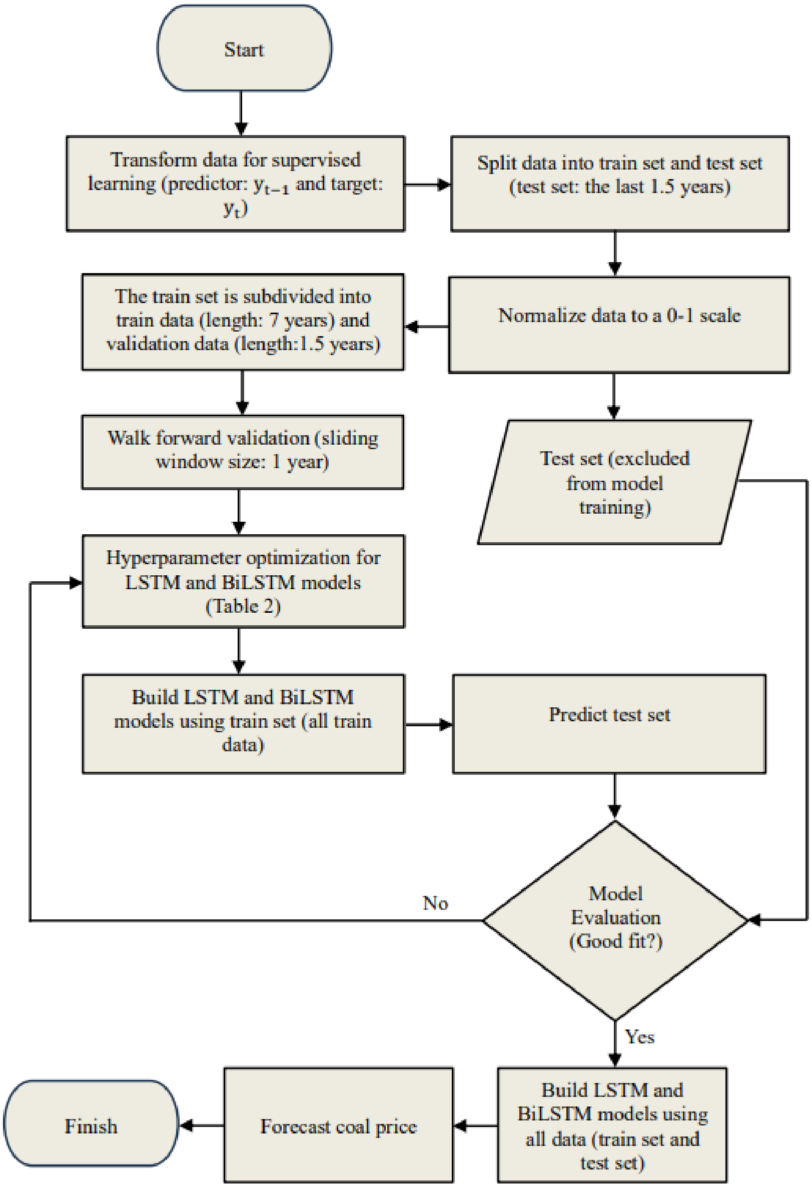


Figure 2. **Flowchart**

2.3 Long Short-Term Memory

Long Short-Term Memory Network (LSTM) is a development of a Recurrent Neural Network designed to overcome the main problem in RNN, which is the loss of gradient on long sequential data [5] [25]. RNN can process sequential data, but there is a weakness when faced with very long sequences. Due to its repetitive nature, RNN tends to shrink exponentially, making the model training process less effective [26].

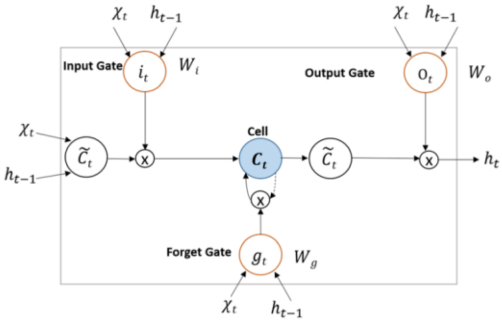


Figure 3. **LSTM Architecture** [25]

LSTMs address the vanishing gradient problem that RNNs face, as seen in **Figure 3**, by enabling the model to learn more complex and lengthy patterns from data. While the basic principle of LSTMs is like that of RNNs, LSTMs have significant structural differences, particularly in their cells, which use multiple gates (input, forget, output) and cell states to retain long-term information. Conventional RNNs have a single layer of neurons with a *tanh* activation function, which helps maintain gradient stability somewhat, but they still struggle with long sequences. In contrast, LSTM cells incorporate additional mechanisms that make them more capable of handling long sequences effectively.

LSTM has a gate mechanism that allows the network to regulate which information should be remembered or forgotten during the learning process. There are three main types of gates in LSTM: the forget gate, the input gate, and the output gate. The forget gate (g_t), as represented in **Equation (2)**, determines which information from the previous cell state $c_{\{t-1\}}$ will be retained or forgotten. The input gate (i_t), described in **Equation (3)**, decides which new information should be added to the cell state. The candidate cell state ($\widetilde{c_t}$), calculated using **Equation (4)**, represents potential updates to the cell state. The current cell state (c_t) ,as shown in **Equation (5)**, combines of the previous cell state and the newly updated information. The output gate (o_t), detailed in **Equation (6)**, controls which part of the information from the current cell state will be used to produce the final output. Finally, the output of the LSTM cell at the current time step h_t is generated as shown in **Equation (7)**, using the output gate and the hyperbolic tangent of the current cell state. This gate mechanism gives the LSTM greater flexibility in managing the flow of information, so it can capture long-term dependencies in sequential data, as detailed in the equations below:

$$g_t = \sigma\left(U_g x_t + W_g h_{t-1} + b_f\right) \tag{2}$$

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i) \tag{3}$$

$$\widetilde{c_t} = tanh(U_x x_t + W_c h_{t-1} + b_c) \tag{4}$$

$$c_t = g_t * c_{t-1} + i_t * \widetilde{c_t} \tag{5}$$

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o) \tag{6}$$

$$h_t = o_t \cdot tanh(c_t) \tag{7}$$

where U and W represent the weight, b denotes the bias term, $tanh$ is the activation function that generates values within the range of -1 to 1, and the symbol $*$ denotes element-wise multiplication.

The main idea of Long Short-Term Memory (LSTM) is that there is a path that connects the previous cell state $c_{\{t-1\}}$ to the current cell state c_t, allowing the information in the cell state to be preserved and passed on to the next cell state with appropriate adjustments. The values of the state cells are designed to contain relevant contextual information from a dataset. Thus, LSTM can retain and utilize long-term information from sequential data, allowing for better learning in tasks like natural language processing and sequence prediction.

2.4 Bidirectional Long Short-Term Memory

Bidirectional Long Short-Term Memory (BiLSTM) is an extension of Long Short-Term Memory (LSTM) that allows neural network models to learn sequential data by considering the previous and next context simultaneously. BiLSTM achieves this by combining two LSTM units that work in opposite directions **[10][27]**.

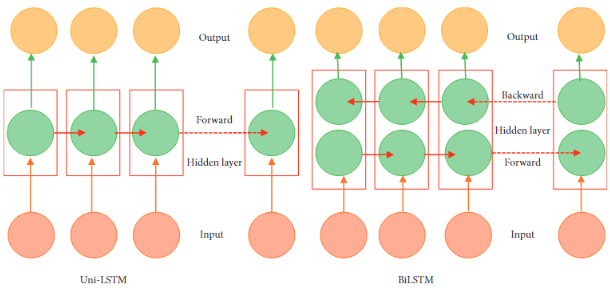


Figure 4. Unidirectional and Bidirectional STM Architecture [10]

Figure 4 shows the architecture of BiLSTM, a combination of two LSTMs in the forward and backward directions. The output of the two-way hidden layer LSTM is as follows:

$$y_t = W_{hy}\overrightarrow{h_t} + W_{hy}\overleftarrow{h_t} \tag{8}$$

The BiLSTM work process involves two separate processing stages using different hidden layers. One of the hidden layers processes the data from the beginning to the end of the sequence, while another processes the data from the end to the beginning. The results from these two hidden layers are then combined to produce a final output that represents a thorough understanding of the given sequence data, as shown in **Equation (8)**. The different direction arrows on $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ indicate that these two hidden layers process data in opposite directions: one from the start to the end (forward), and the other from the end to the beginning (backward) with W_{hy} representing the weight in hidden states. BiLSTM sometimes has advantages over conventional LSTM, especially in terms of accuracy in sequential data processing.

2.5 Walk-Forward Validation

Walk-forward validation is a technique often used in backtesting and cross-validation processes, especially in the context of time series analysis. It allows the performance evaluation of a strategy or a model using historical data without that much extensive retraining. The main concept involves continuous sliding windows inside and outside the data sample.

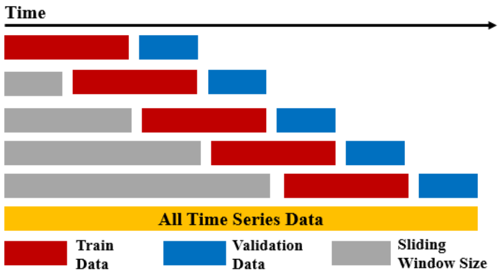


Figure 5. Walk Forward Validation Scenario

In walk forward validation, we use a constant sliding window to cover the entire training and test data. This window has a fixed size for both types of data (training and test) and consistent shifts. With this approach, we can incrementally create multiple training and test data pairs to better evaluate the strategy or model under test [28]. The process starts by using the first part of the historical data as training data to train an initial model. Then, the model is tested using the next part of the historical data as test data. Once the test is complete, the process is repeated by moving the sliding window forward. In this way, we can measure how well the model or strategy can adapt to gradually changing conditions, resulting in a more realistic and informative evaluation of the actual performance of the model or strategy under test.

2.6 Model Evaluation

Forecasting accuracy is determined by assessing how well the model performs on a new dataset omitted in model building [29]. Accuracy measures commonly used for forecasting include mean absolute percentage error (MAPE) and root mean square error (RMSE). MAPE is an accuracy measure that calculates the difference between actual and forecast data in percentage form [30]. A lower MAPE value indicates more accurate forecasting, as it shows a smaller average percentage deviation from the actual values. RMSE is the square root of the average of the squared differences between predicted values and actual values. The MAPE and RMSE values are calculated using **Equation (9)** and **Equation (10)**.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y_t}}{y_t}\right| \times 100\% \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y_t})^2} \tag{10}$$

Description:

(n) : the number of observations

(y_t) : the actual value at time (t)

$(\hat{y}_t)$: the predicted value at time (t)

Table 1. Interpretation of MAPE for Forecasting Accuracy (Lewis, 1982)

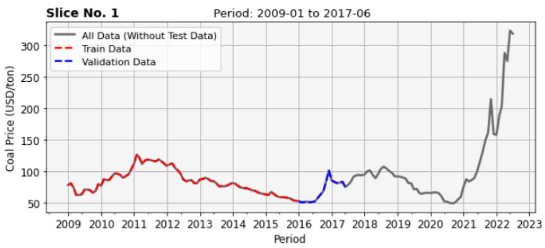
MAPE value	Accuracy of forecast
≤ 10 %	Highly Accurate
11 – 20 %	Good
21 – 50 %	Reasonable
≥ 51 %	Inaccurate

Selecting an appropriate evaluation metric is crucial in assessing the performance of a forecasting model. MAPE and RMSE provide different information and can be used complementarily. MAPE measures the error as a percentage, making it easy to interpret and giving a direct idea of the forecast deviation from the actual value. Meanwhile, RMSE emphasizes the magnitude of individual errors, giving greater weight to larger errors. A lower RMSE and MAPE value indicates more accurate forecasting.

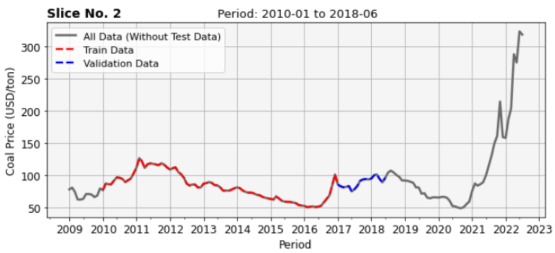
3. RESULTS AND DISCUSSION

3.1 Walk-Forward Validation Scenario

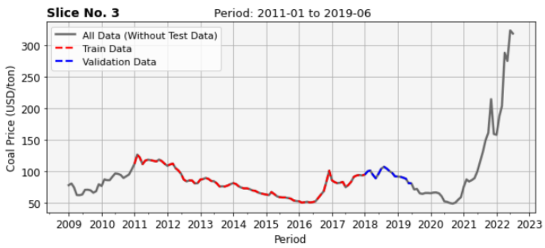
Six data segments are acquired in the walk-forward validation scenario, as seen in **Figure 6**, with training data length, validation data length, and sliding window width of 7 years, 1.5 years, and one year, respectively. In **Figure 6**, the training set is partitioned into training and validation data and then shifted by one year with a constant size. This procedure continues until the most recent data is incorporated. Each data segment undergoes hyperparameter tuning from 90 available combinations in **Table 2**. Every hyperparameter combination is applied in constructing LSTM and BiLSTM models using the training data, and then the models' performance is assessed on the validation data. In this scenario, there are a total of 540 RMSE and MAPE values, with each hyperparameter combination having six RMSE and MAPE values, respectively. The best models from LSTM and BiLSTM are selected based on the criteria of the smallest average RMSE and average MAPE values between hyperparameter combinations.



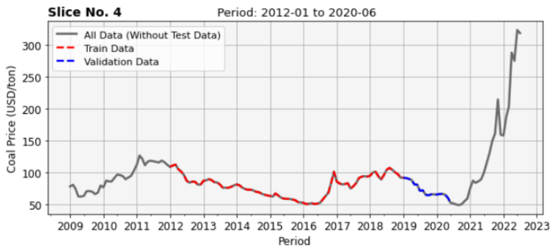
(a)



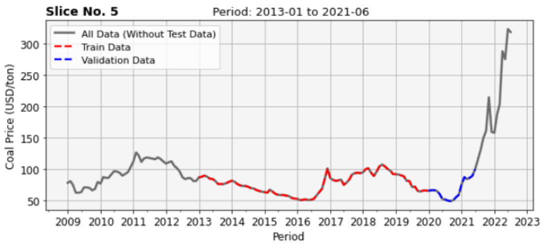
(b)



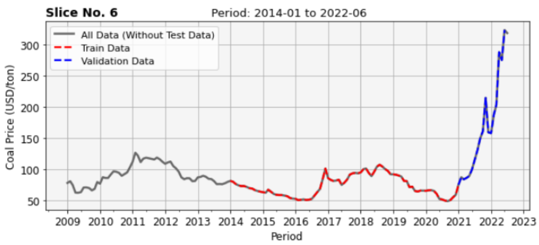
(c)



(d)



(e)



(f)

Figure 6. Walk Forward Validation Scenario Using Training Set (Training and Validation Data) (a) – (f) Data Slice No 1 - 6

Table 2. Hyperparameters of LSTM and BiLSTM

Hyperparameter	Value
Optimizer	Adaptive Moment Estimation (Adam), Root Mean Square Propagation (RMSProp)
Neuron	20, 32, 64, 100, 200
Learning Rate	0.01, 0.001, 0.0001
Epoch	100, 200, 500
Layer	1
Dropout	0.2
Activation Function	Rectified Linear Unit (ReLU)

Optimizer is essential for adjusting model parameters during training to improve performance and convergence. This study evaluates two optimizers, Adam and RMSProp, in the context of hyperparameter tuning. The RMSprop optimizer reduces fluctuations during training and enhances the model's speed and accuracy. It improves performance by updating model parameters, such as weights and biases [31]. In contrast, Adam offers several advantages, including invariant parameter updates to gradient rescaling, step sizes that are approximately bounded by the step size hyperparameter, no need for a stationary objective, compatibility with sparse gradients, and inherent step size annealing [32]. We also apply dropout 0.2 to prevent overfitting and make more efficient computation, with 20% of neurons randomly dropped in the training process. For the activation function that allows the neural network to approximate non-linear functions and solve more complex tasks, this work implements the ReLU because of its computational simplicity and representational sparsity [33].

3.2 Model Evaluation and Forecasting

To provide the best model of the LSTM and BiLSTM, hyperparameter tuning is performed using a grid search method, which explored 90 potential hyperparameter combinations, as detailed in Table 2. The optimal configurations for LSTM and BiLSTM models are shown in Table 3. Mathematically, the BiLSTM model demonstrates superior performance compared to LSTM, as indicated by its lower average RMSE and MAPE values on the validation dataset.

Table 3. The Best Models

Model	Hyperparameter				Average RMSE of Validation Data	Average MAPE of Validation Data
	Optimizer	Neuron	Learning Rate	Epoch		
LSTM	Adam	200	0.01	500	11.993	10.442
BiLSTM	Adam	100	0.001	500	10.485	7.847

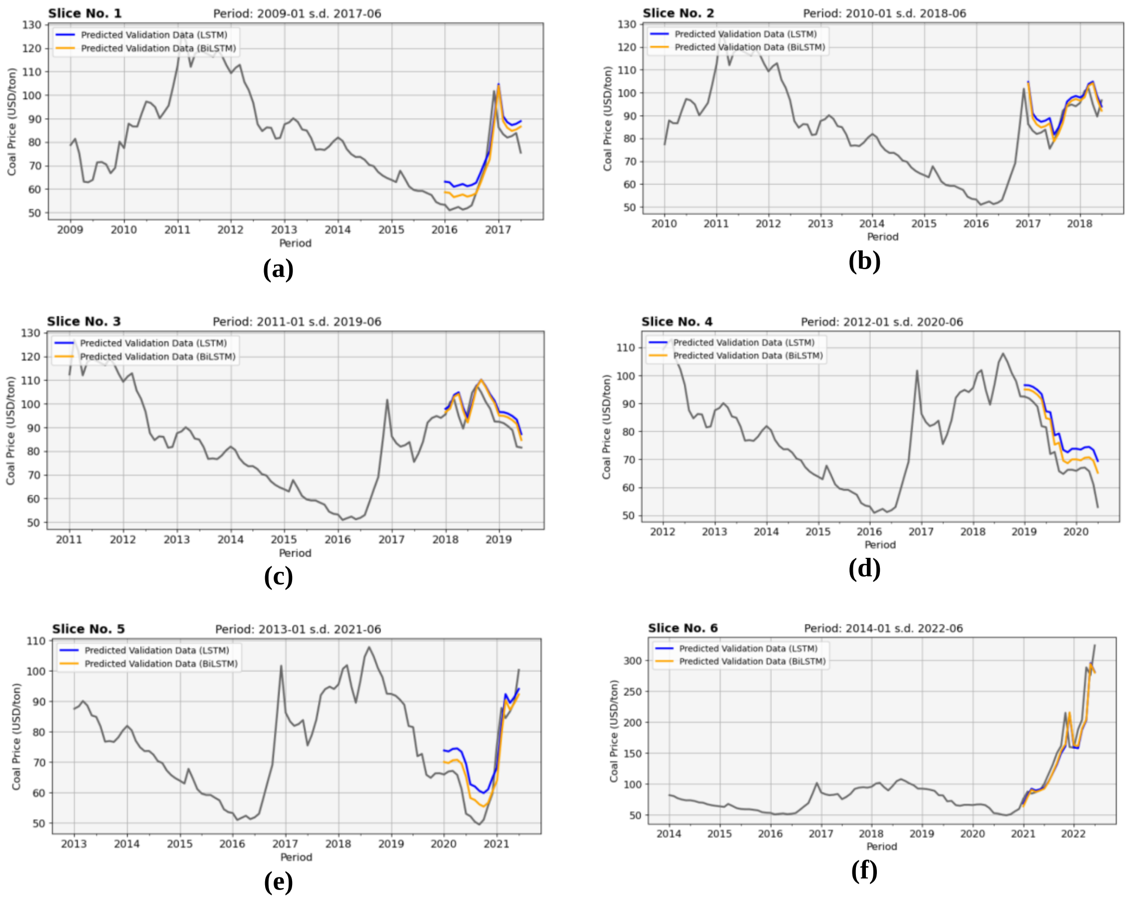
The Diebold-Mariano (DM) test is applied to assess whether there is a significant difference between the two forecasts [21]. The DM test results in Table 4 provide a comparative analysis of forecasting accuracy between the LSTM and BiLSTM models based on validation data across multiple segments. This study utilizes squared error (SE) and absolute error (AE) as loss functions for a more robust analysis. For most segments, the DM-test statistic is negative, indicating that the BiLSTM model generally outperforms the LSTM model regarding forecasting accuracy. It is supported by the *p*-values associated with the DM-test, where values less than 0.01 (denoted by a) and less than 0.10 (denoted by b) suggest a statistically significant difference in favor of the BiLSTM model at the 1% and 10% significance levels, respectively. For segments 1 to 5, both SE and AE show that BiLSTM is more accurate than the LSTM model at a 1% significance level. However, segment 6 shows a *p*-value of 0.064 under the SE loss function, indicating marginal significance at the 10% level. In contrast, the AE loss function does not indicate a significant difference for this segment. These findings highlight that the BiLSTM model consistently demonstrates superior predictive performance over the LSTM model across multiple segments, with the differences being statistically significant in most cases.

Table 4. Model Evaluation (Validation Data)

Segment (Slice)	MAPE		RMSE		Diebold-Mariano Test			
					Loss Function			
					SE		AE	
	LSTM	BiLSTM	LSTM	BiLSTM	DM-Test	*p*-value	DM-Test	*p*-value
1	13.24	8.74	9.44	7.68	-3.037	0.001[a]	-4.687	0.0[a]
2	6.32	5.22	7.11	6.31	-2.704	0.003[a]	-2.557	0.005[a]
3	5.8	4.99	6.11	5.35	-2.977	0.001[a]	-2.557	0.005[a]
4	12.4	7.96	9.21	6.4	-6.761	0.0[a]	-13.391	0.0[a]
5	13.08	8.46	8.69	6.47	-2.938	0.002[a]	-4.034	0.0[a]
6	11.81	11.71	31.4	30.7	-1.525	0.064[b]	-0.869	0.193
Average	10.442	7.847	11.993	10.485				

Note: [a] = 1% significance level; [b] = 10% significance level; SE: Squared Error; AE; Absolute Error

Figure 7 presents the validation data prediction outcomes generated by the top-performing LSTM and BiLSTM models in each data segment within the walk-forward validation scenario. Visually, both models produce excellent predictions due to their ability to track the upward and downward trends in actual coal prices. Nevertheless, a distinct observation emerges: the validation data prediction derived from the BiLSTM model tends to align more closely with the actual validation data across all segments.



(a) (b) (c) (d) (e) (f)

Figure 7. Validation Data Prediction in Walk Forward Validation Scenario (a) – (f) Data Slice No 1 - 6

Furthermore, we rebuild both models using the entire training set according to the hyperparameters listed in **Table 3**. The testing data (spanning the most recent 1.5 years) excluded from the walk-forward validation scenario is forecasted using these models. Our objective at this stage is to ensure that the constructed models exhibit consistent performance when predicting unknown real-world data. Despite the extreme increase observed in the historical data from two years earlier (2021–2022), it is evident that these two models excel in identifying the downward trend in coal prices in the testing data (**Figure 8**). The prediction results on the testing data further confirm that BiLSTM surpasses LSTM, as evidenced by lower MAPE and RMSE values, aligning with earlier findings. As shown in **Table 5**, the DM test demonstrates that BiLSTM outperforms LSTM, with statistically significant results based on SE and AE at the 5% and 10% significance levels, respectively. The bidirectional information from both past (backward) and future (forward) contexts, processed simultaneously in the BiLSTM algorithm, performs well on this data. It contributes to demonstrating the excellent performance of BiLSTM compared to LSTM. Therefore, this study on coal price forecasting aligns with several previous studies **[10][12] [13]** but differs from the research on air pollution index forecasting, which found no significant difference between LSTM and BiLSTM **[14]**.

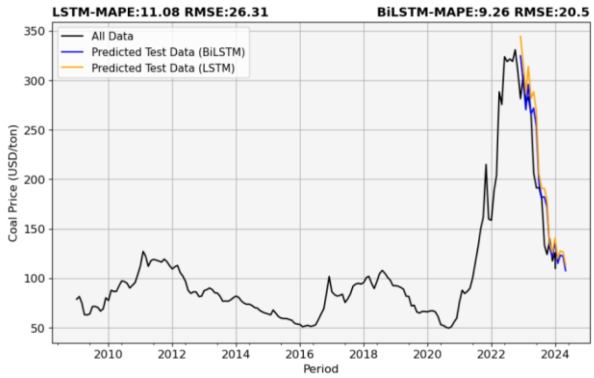


Figure 8. Testing Data Prediction from LSTM and BiLSTM Models

Table 5. Diebold-Mariano Test (Testing Data)

Loss Function			
SE		AE	
DM-Test	***p*-value**	**DM-Test**	***p*-value**
-1.964	0.025[a]	-1.506	0.066[b]

Note: [a] = 5% significance level; [b] = 10% significance level; SE: Squared Error; AE; Absolute Error

In a prior study, Prahesti et al. (2022) attempted to forecast coal prices using the same data as this work. It obtained MAPE values of 23.14 and 17.66, respectively, on the testing dataset using ARIMA and Transfer Function models with a period of data from January 2015 to July 2022 **[15]**. A rough comparison of current work with the previous study reveals a significant performance obtained from the LSTM and BiLSTM models. The MAPE value of the BiLSTM model is approximately two times smaller than that of the previous study. Nevertheless, we also speculate that the larger dataset contributes to better results in this work.

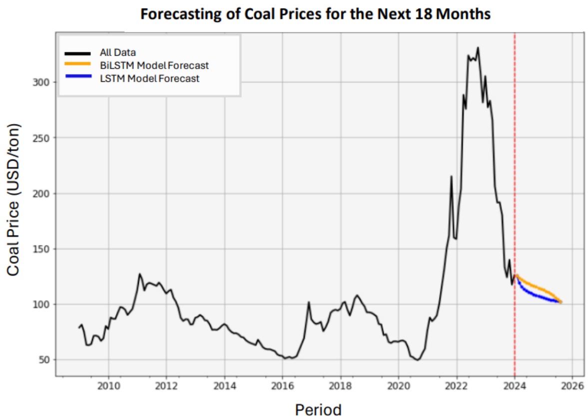


Figure 9. Forecasting of Coal Prices

Table 6. Detailed Forecasting of Coal Prices

Period	Coal Price (USD/ton)		Period	Coal Price (USD/ton)	
	LSTM	BiLSTM		LSTM	BiLSTM
Feb-24	119.27	123.69	Nov-24	106.29	113.56
Mar-24	115.68	122.03	Dec-24	105.63	112.58
Apr-24	113.54	120.64	Jan-25	105.01	111.53
May-24	112.02	119.42	Feb-25	104.44	110.39
Jun-24	110.77	118.30	Mar-25	103.91	109.12
Jul-24	109.69	117.28	Apr-25	103.42	107.66
Aug-24	108.72	116.34	May-25	102.96	105.96
Sep-24	107.83	115.42	Jun-25	102.53	104.02
Oct-24	107.02	114.50	Jul-25	102.13	101.93

In the final stage, we utilize all available data (training, validation, and testing data) to construct the final model for forecasting coal prices over the next 18 months. Despite the best performance of the BiLSTM model compared to LSTM in this case, we still present the forecasting result obtained from LSTM, as depicted in **Figure 9** and **Table 6**. The reliable forecasting result is the BiLSTM model because of its performance from the previous evaluation model, and the LSTM model result is simply for comparison. The forecasts from both models show considerable similarity, indicating a close to linear downward trend in coal prices over the next 18 months. These forecasting results align with those of the International Energy Agency (2023), which also projected a decline in coal demand beginning from early 2024 through to 2026 **[13]**. In Europe and the United States, the growth of clean energy has led to a structural decline in coal use. While global coal demand has experienced declines in the past, these were typically brief and driven by extraordinary events such as the collapse of the Soviet Union or the Covid-19 crisis. However, the current decline appears more structural, driven by the substantial and sustained expansion of clean energy technologies **[2][3]**.

4. CONCLUSIONS

In the walk-forward validation scenario, the BiLSTM model consistently outperforms the LSTM model based on average RMSE and MAPE values. Specifically, the BiLSTM achieves an average MAPE of 7.847 and an average RMSE of 10.485, compared to 10.442 and 11.993 for LSTM, respectively. The Diebold-Mariano (DM) test with SE and AR loss functions further supports this finding, with most segments showing significant improvements in favor of BiLSTM, evidenced by negative DM-test statistics and p-values below 0.01 or 0.10. This superior performance extends to the testing phase, where BiLSTM continues to show lower MAPE and RMSE values and a significant DM-test. Ultimately, while both models forecast a similar downward trend over the next 18 months, the BiLSTM's enhanced accuracy, reflected in lower error metrics and statistical significance, highlights its robustness and reliability for forecasting. This conclusion aligns with broader industry projections indicating a structural decline in coal demand driven by clean energy advancements. Future research could enhance this analysis by incorporating intervention analysis to explain the effects of external factors on coal prices from 2022 to 2023 and utilize other explanatory variables.

ACKNOWLEDGMENT

We gratefully acknowledge the Ministry of Energy and Mineral Resources (KESDM) for providing the reference coal price data and the Department of Statistics and Data Science for funding this research.

REFERENCES

[1] S. Makkl, “Analis Ungkap Alasan Konflik Rusia-Ukraina Bisa Terbangkan Batu Bara,” CNN Indonesia. Accessed: Mar. 12, 2024. [Online]. Available: https://www.cnnindonesia.com/ekonomi/20220307174517-85-767904/analis-ungkap-alasan-konflik-rusia-ukraina-bisa-terbangkan-batu-bara.

[2] IEA, “Global coal demand set to remain at record levels in 2023,” International Energy Agency’s. Accessed: Aug. 21, 2024. [Online]. Available: https://www.iea.org/news/global-coal-demand-set-to-remain-at-record-levels-in-2023.

[3] IEA, “Global coal demand expected to decline in coming years,” International Energy Agency’s. Accessed: Aug. 21, 2024. [Online]. Available: https://www.iea.org/news/global-coal-demand-expected-to-decline-in-coming-years.

[4] ESDM, “HARGA BATUBARA ACUAN,” KEMENTERIAN ENERGI DAN SUMBER DAYA MINERAL. Accessed: Mar. 12, 2024. [Online]. Available: https://www.minerba.esdm.go.id/harga_acuan.

[5] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[6] T. Gao, Y. Chai, and Y. Liu, “Applying Long Short Term Momory Neural Networks for Predicting Stock Closing Price,” in 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), IEEE, Nov. 2017, pp. 575–578. doi: 10.1109/ICSESS.2017.8342981.

[7] Z. Zhang, “Comparison of LSTM and ARIMA in Price Forecasting: Evidence from Five Indexes,” 2024, pp. 40–46. doi: 10.2991/978-94-6463-268-2_6.

[8] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, “A Comparison of ARIMA and LSTM in Forecasting Time Series,” in Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, Institute of Electrical and Electronics Engineers Inc., Jul. 2018, pp. 1394–1401. doi: 10.1109/ICMLA.2018.00227.

[9] H. Wu, S. Chen, and Y. Ding, “Comparison of ARIMA and LSTM for Stock Price Prediction,” Financial Engineering and Risk Management, vol. 6, no. 1, 2023, doi: 10.23977/ferm.2023.060101.

[10] R. L. Abduljabbar, H. Dia, and P.-W. Tsai, “Unidirectional and Bidirectional LSTM Models for Short-Term Traffic Prediction,” *J Adv Transp*, vol. 2021, pp. 1–16, Mar. 2021, doi: 10.1155/2021/5589075.

[11] S. Wang, “A Stock Price Prediction Method Based on BiLSTM and Improved Transformer,” *IEEE Access*, vol. 11, pp. 104211–104223, 2023, doi: 10.1109/ACCESS.2023.3296308.

[12] S. Siami-Namini, N. Tavakoli, and A. S. Namin, “The Performance of LSTM and BiLSTM in Forecasting Time Series,” in 2019 IEEE International Conference on Big Data (Big Data), IEEE, Dec. 2019, pp. 3285–3292. doi: 10.1109/BigData47090.2019.9005997.

[13] D. G. da Silva and A. A. de M. Meneses, “Comparing Long Short-Term Memory (LSTM) and bidirectional LSTM deep neural networks for power consumption prediction,” Energy Reports, vol. 10, pp. 3315–3334, Nov. 2023, doi: 10.1016/j.egyr.2023.09.175.

[14] T. Handhayani, I. Lewenusa, D. E. Herwindiati, and J. Hendryli, “A Comparison of LSTM and BiLSTM for Forecasting the Air Pollution Index and Meteorological Conditions in Jakarta,” in 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), IEEE, Dec. 2022, pp. 334–339. doi: 10.1109/ISRITI56927.2022.10053078.

[15] S. P. Prahesti, I. D. Sulvianti, and Y. Angraini, “PERAMALAN HARGA BATU BARA ACUAN MENGGUNAKAN METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE DAN FUNGSI TRANSFER,” *Xplore: Journal of Statistics*, vol. 12, no. 1, pp. 1–11, Jan. 2023, doi: 10.29244/xplore.v12i1.1100.

[16] Y. Zheng, K. Yuan, and L. Dai, “Research on Coal Price Forecast based on ARIMA and SVM combination Model,” in *E3S Web of Conferences*, EDP Sciences, May 2021. doi: 10.1051/e3sconf/202125702008.

[17] X. Liu, “Research on the forecast of coal price based on LSTM with improved Adam optimizer,” *J Phys Conf Ser*, vol. 1941, no. 1, p. 012069, Jun. 2021, doi: 10.1088/1742-6596/1941/1/012069.

[18] X. Xu and Y. Zhang, “Thermal coal price forecasting via the neural network,” *Intelligent Systems with Applications*, vol. 14, p. 84, 2022, doi: 10.1016/j.iswa.2022.20.

[19] Z. Alameer, A. Fathalla, K. Li, H. Ye, and Z. Jianhua, “Multistep-ahead forecasting of coal prices using a hybrid deep learning model,” *Resources Policy*, vol. 65, p. 101588, Mar. 2020, doi: 10.1016/j.resourpol.2020.101588.

[20] S. Pan, “Coal Price Prediction based on LSTM,” *Journal of Physics: Conference Series*, vol. 1802, no. 4, p. 042055, Mar. 2021, doi: 10.1088/1742-6596/1802/4/042055.

[21] F.X. Diebold and R.S. Mariano, “Comparing predictive accuracy,” *JBES*, vol. 13, pp. 253-263, Jul 1995.

[22] S. Panigrahi and H. S. Behera,” Effect of Normalization Techniques on Univariate Time Series Forecasting using Evolutionary Higher Order Neural Network,” *IJEAT*, vol.3, Dec 2013.

[23] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance,” *Appl Soft Comput*, vol.97, 2020, doi: 10.1016/j.asoc.2019.105524.

[24] J. N. Van Rijn and F. Hutter,”Hyperparameter importance across datasets,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Jul. 2018, pp. 2367–2376. doi: 10.1145/3219819.3220058.

[25] A. García-Medina and E. Aguayo-Moreno, “LSTM–GARCH Hybrid Model for the Prediction of Volatility in Cryptocurrency Portfolios,” *Comput Econ*, vol. 63, no. 4, pp. 1511–1542, Apr. 2024, doi: 10.1007/s10614-023-10373-8.

[26] L. Mualifah, K. A. Notodipuro, and A. M. Soleh, “A comparative study of GARCH, LSTM, and HYBRID GARCH– LSTM models for analyzing the dynamical pattern of stock price volatility,” M.S thesis, IPB University, Bogor, 2022.

[27] D. A. Dzulhijjah, Kusrini, and K. A. Yuana, “IMPLEMENTASI ALGORITMA BI-LSTM DALAM MENDETEKSI ENTITAS WAKTU DAN LOKASI KEBAKARAN HUTAN,” 2023.

[28] R. Pardo, “The Evaluation and Optimization of Trading Strategies Second Edition,” 2007.

[29] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne: OTexts, 2021.

[30] C.D. Lewis, *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*, London; Boston: Butterworth Scientific, 1982.

[31] P. Nimmani, S. Vodithala, and V. Polepally, “Neural Network Based Integrated Model for Information Retrieval,” in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, May 2021, pp. 1286–1289. doi: 10.1109/ICICCS51141.2021.9432241.

[32] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980.

[33] X. Glorot, A. Bordes, and Y. Bengio, “Deep Sparse Rectifier Neural Networks,” 2011.