Vol. 7, No. 4, December 2025

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Oil and Gas Production Forecasting Based on LSTM Model: A Case Study of PT Pertamina Hulu Rokan Zone 4

Angel Caroline Billan[1], Usman Ependi[2]

[1,2]Informatics Department, Postgraduate Program, Bina Darma University, Palembang, Indonesia

Email: 242420010@students.binadarma.ac.id[1], u.ependi@binadarma.ac.id[2]

Received: Sept 30, 2025

Revised: October 11, 2025

Accepted: Nov 21, 2025

Published: Dec 15, 2025

Corresponding Author:

Author Name*:

Usman Ependi

Email*:

u.ependi@binadarma.ac.id

DOI:

10.63158/journalisi.v7i4.1285

Abstract. This study addresses the critical need for accurate oil and gas production forecasting to support strategic decision-making in Indonesia's energy sector. PT Pertamina Hulu Rokan Zone 4 (PHR Zona 4), a key player in national energy production, frequently encounters technical and external operational challenges. To tackle these issues, this research proposes a deep learning-based predictive model using the Long Short-Term Memory (LSTM) architecture, structured in an encoder-decoder format and enhanced with an attention mechanism. The model was trained and tested on historical oil and gas production data from PHR Zona 4, evaluated under two data-splitting scenarios: 80:20 and 90:10. Model performance was assessed using Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Results from the 80:20 scenario showed RMSE of 5.83, MAE of 5.54, MAPE of 1.71%, and R^2 of -1.97, suggesting difficulties in capturing extreme data fluctuations. However, the 90:10 scenario demonstrated significantly improved performance with RMSE of 0.42, MAE of 0.36, MAPE of 0.11%, and R^2 of 0.00, indicating better trend prediction stability. The novelty of this study lies in the integration of attention mechanisms within the LSTM encoder-decoder framework for oil and gas time series forecasting, offering enhanced accuracy and robustness. This research provides a valuable foundation for future improvements in predictive analytics and operational efficiency in the oil and gas industry.

Keywords: Attention Mechanism, LSTM, MAPE, Oil and Gas Production Forecasting, RMSE

1. INTRODUCTION

Oil and natural gas are vital resources that play a strategic role in maintaining energy security, both nationally and globally. Amid rising global energy demands and pressure to achieve a sustainable energy transition, the upstream oil and gas sector becomes a key determinant of future energy policies. Indonesia, blessed with abundant natural resources, plays a significant role through its state-owned enterprises, particularly PT Pertamina (Persero) and its subsidiaries. One of the key entities in this sector is PT Pertamina Hulu Rokan (PHR) Zone 4, which consistently engages in exploration and production (E&P) activities to support national oil and gas production targets.

PHR Zone 4 has a significant production profile. To date, its cumulative oil production has reached 333.17 million barrels (MMSTB), with remaining reserves of approximately 20,432 MSTB. The daily oil production averages around 3,716 BOPD, while gas production reaches 7.90 million standard cubic feet per day (MMSCFD). Despite these strong figures, meeting production targets is not always straightforward. The company faces numerous challenges, ranging from technical and operational issues to external factors such as market volatility and regulatory policy changes.

In this context, the role of SKK Migas as the regulator of the upstream oil and gas sector is crucial. SKK Migas sets mandatory production targets for each working area, including Zone 4. To ensure these targets are met, accurate and responsive production forecasting capabilities are needed. Precise forecasts serve as the foundation for developing operational strategies and making more effective decisions.

The core problem addressed in this study is how to build a forecasting system capable of accurately projecting oil and gas production within PHR Zone 4's operational environment. This is where artificial intelligence (AI) plays a vital role. One prominent method for handling time-series data is the Long Short-Term Memory (LSTM) architecture, a type of deep learning model that belongs to the Recurrent Neural Network (RNN) family.

LSTM excels in recognizing long-term patterns and retaining important information over relevant time periods. This capability makes it especially suitable for processing oil and gas production data, which tends to be volatile, dynamic, and influenced by various technical and external factors. Therefore, this study aims to develop an LSTM-based forecasting model using historical oil and gas production data from PHR Zone 4. The model's performance is evaluated using two key metrics: Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

Previous studies have demonstrated the effectiveness of LSTM in handling time-series data across different sectors. For example, research using LSTM for sales forecasting showed high accuracy based on repeated regression testing. Wiranda (2019) also reported the best results in pharmaceutical product sales forecasting using MAPE and RMSE metrics in the fifth trial. In the energy sector, LSTM has been used to forecast global oil prices and successfully capture price fluctuation patterns. Furthermore, this method has proven effective in predicting cryptocurrency prices and even classifying crowdsourced data.

The novelty of this study lies in the direct application of the LSTM algorithm to oil and gas production data from Indonesia, specifically within PHR Zone 4—a relatively unexplored approach in Pertamina's operational context. Beyond the technical contribution, this research also supports Pertamina's digital transformation agenda and the implementation of Industry 4.0 principles. Moreover, the forecasting model developed in this study is expected to be integrated in real-time into the company's decision-making systems, paving the way for automated production monitoring and greater operational efficiency.

2. METHODS

This research was conducted using an applied technology approach based on data science and artificial intelligence, designed systematically and step-by-step to develop a predictive model for oil and gas production using the Long Short-Term Memory (LSTM) algorithm. This approach is highly relevant in the context of the modern oil and gas industry, which relies on historical production data as a foundation for strategic decision-making particularly in lifting planning, well optimization, and operational efficiency. The research stages are organized in a multi-level process flow as shown in Figure 2, consisting of three main blocks: input, process, and output. Each phase is designed to be

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

interconnected, resulting in a predictive system that is applicable within the operational environment of PT Pertamina Hulu Rokan Zone 4.

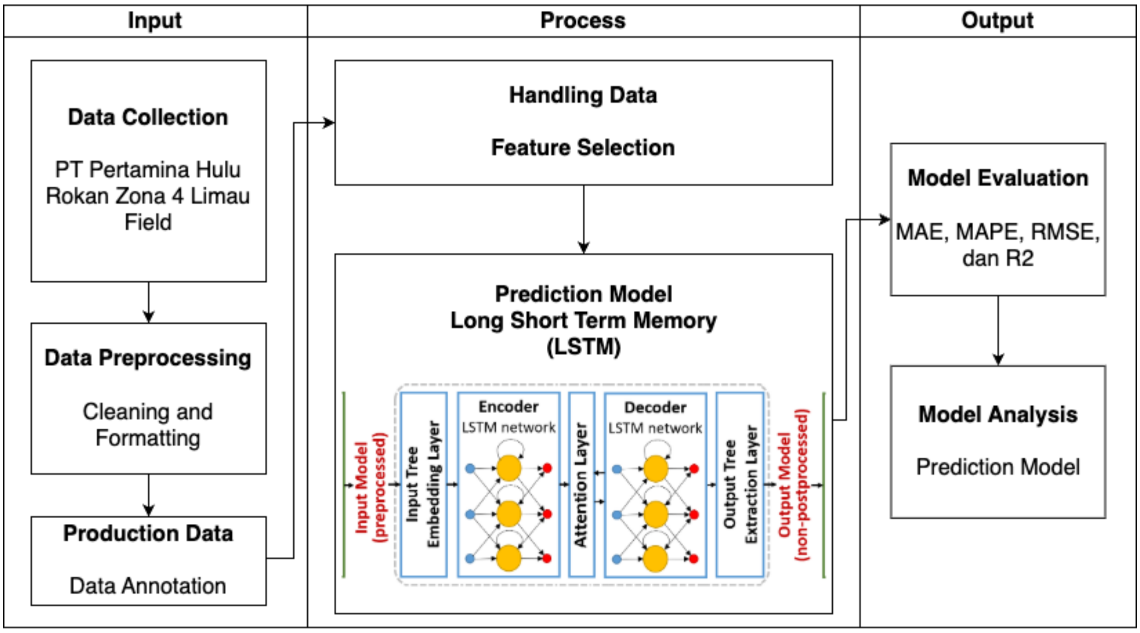

Figure 1. Proposed Research Workflow

2.1. Data Preparation (Input)

Figure 1 illustrates the development flow of the oil and gas production prediction model based on the Long Short-Term Memory (LSTM) algorithm. This phase begins with the collection of data from PT Pertamina Hulu Rokan Zone 4, Limau Field, specifically daily oil and gas production data. The dataset includes five months of historical production data, covering essential parameters such as oil production volume (BOPD), gas production (MMSCFD), reservoir pressure, and other relevant operational technical data. Once collected, the data undergoes preprocessing, including cleaning of extreme values, removal of outliers, filling in missing values, and formatting adjustments to ensure compatibility with the deep learning model. Further, the data is annotated based on time periods and production characteristics, and separated into oil and gas datasets. This is then split into training and testing sets, maintaining balanced data distribution. Table 1 and Table 2 show sample datasets used for developing the oil and gas prediction models.

Table 1. Sample Oil Production Data (BEL-22)

Period	BEL-22 Gross	Wc	Net
01-Jan-25	0	0	0
02-Jan-25	0	0	0
03-Jan-25	0	0	0
04-Jan-25	0	0	0
05-Jan-25	0	0	0
...	...	...	...
26-Jun-25	476	98	10
27-Jun-25	476	98	10
28-Jun-25	465	98	9
29-Jun-25	465	98	9
30-Jun-25	465	98	9

Table 2. Sample Gas Production Data

Period	BEL-22	25	26	28	29	30	34	37	40
01-Jan-25	0	0	0	0	0	0	0	0	79.8
02-Jan-25	0	0	0	0	0	0	0	0	79.8
03-Jan-25	0	0	0	0	0	0	0	0	79.8
04-Jan-25	0	0	0	0	0	0	0	0	79.8
05-Jan-25	0	0	0	0	0	0	0	0	79.8
...	...	...	...	...	...	...	...	...	...
26-Jun-25	0	0	0	0	0	85	0	120	120
27-Jun-25	0	0	0	0	0	85	0	120	120
28-Jun-25	0	0	0	0	0	85	0	120	120
29-Jun-25	0	0	0	0	0	85	0	120	120
30-Jun-25	0	0	0	0	0	85	0	120	120

2.2. Model Development (Process)

Once the oil and gas production data from PT Pertamina Hulu Rokan Zone 4 is prepared, the next step is model training, starting with feature selection. This step ensures the model can handle high variability in production data without introducing bias toward

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

specific production patterns. Therefore, selecting the right features is critical to capture relevant trends and minimize noise in the data.

The prediction model used in this study is Long Short-Term Memory (LSTM) with an encoder-decoder architecture, enhanced by an attention mechanism. This LSTM architecture enables the model to understand complex and fluctuating patterns in time-series data. The encoder component captures the sequential patterns from historical data, while the attention layer helps the model focus on the most relevant information for prediction. The attention mechanism assigns different weights to each time step in the input sequence, improving the model's prediction accuracy. Mathematical Representation of LSTM as shown in Equation 1.

$$
\begin{gathered}
i_t = \sigma\left(W_i \cdot \left[h_{\{t-1\}}, x_t\right] + b_i\right) \\
f_t = \sigma\left(W_f \cdot \left[h_{\{t-1\}}, x_t\right] + b_f\right) \\
o_t = \sigma\left(W_o \cdot \left[h_{\{t-1\}}, x_t\right] + b_o\right) \\
c_t = f_t \cdot c_{\{t-1\}} + i_t \cdot tanh\left(W_c \cdot \left[h_{\{t-1\}}, x_t\right] + b_c\right) \\
h_t = o_t \cdot \tanh(c_t)
\end{gathered} \tag{1}
$$

Where:

i_t = input gate

f_t = forget gate

o_t = output gate

c_t = cell state

h_t = hidden state

W_i, Wf, Wo, Wc = weight matrices

b_i, bf, bo, bc = bias terms

The attention mechanism in the LSTM model assigns attention scores to certain time steps based on their relevance to the current prediction, enhancing the model's ability to focus on critical features in the input sequence. Attention Mechanism Formula as shown in Equation 2.

$$
\begin{gathered}
\alpha\_t = \text{softmax}(W\_a \cdot h\_t + b\_a) \\
\hat{h}\_t = \alpha\_t \cdot h\_t
\end{gathered} \tag{2}
$$

Where:

α_t = attention weight at time step t

W_a = attention weight matrix

$\hat{h}_t$ = weighted hidden state

In the architecture shown in Figure 1, preprocessed input data passes through an embedding layer, followed by an LSTM encoder. The output from the encoder is enhanced by the attention mechanism, allowing the model to assign varying weights to different time steps. This refined information is then processed by the LSTM decoder to generate production volume forecasts for future periods.

2.3. Model Evaluation (Output)

The final output of this process is a trained prediction model, which is evaluated using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) as the performance metrics. The evaluation is done by comparing the model's forecast results with the actual production data to assess accuracy and reliability. In addition, a performance analysis is conducted to identify the model's strengths and weaknesses in handling real-world production data. This stage is crucial to ensure that the developed model is truly feasible for implementation in the operational system of PT Pertamina Hulu Rokan Zone 4 and can support precise and adaptive data-driven decision-making.

3. RESULTS AND DISCUSSION

3.1. Performance of the Gas Prediction Model

In this experiment, a Long Short-Term Memory (LSTM) model with an encoder-decoder architecture, enhanced by an attention mechanism, was applied to predict oil and gas production at PT Pertamina Hulu Rokan Zone 4. The data was divided into two training scenarios: 80:20 and 90:10.

1) 80:20 Training Scenario

In the training scenario using an 80:20 data split, the training results showed a significant decline in training loss, as depicted in the loss curve in Figure 2. The MSE value for the validation data also showed a downward trend, although there were some fluctuations in the early stages of training. This indicates that the model was able to learn well from the training data; however, there were signs of overfitting at the beginning, where the model became more tuned to the training data than the validation data.

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

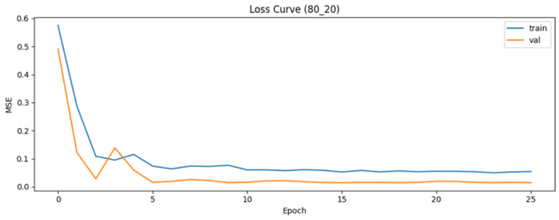

Figure 2. Training and Validation Loss Values

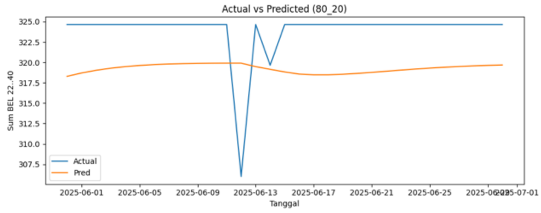

Figure 3. Comparison of Predictions and Validation Data

The comparison between actual and predicted values on the validation set, as shown in Figure 3, demonstrates that while the model could generally follow the overall trend of the production data, some data points showed significant prediction errors—particularly around the date 2025-06-13. Large fluctuations in the residuals, as seen in Figure 4, confirm that the model struggled to predict accurately at those specific points. These errors may have been caused by sudden changes in production data or undetected external factors.

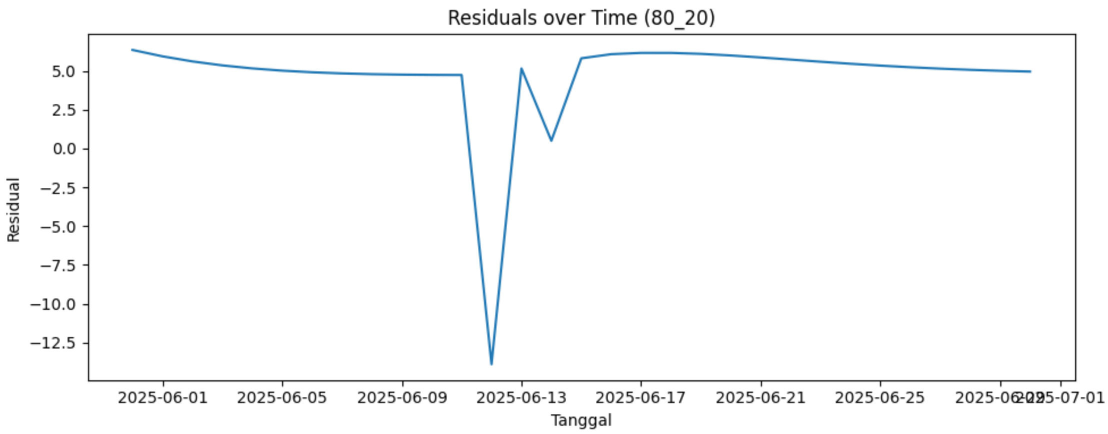

Figure 4. Residual Fluctuations

The residual distribution, shown in the histogram in Figure 5, reveals that most prediction errors clustered around zero, with a few outliers on the right-hand side, indicating overestimations compared to actual values. This is further supported by the QQ-plot of residuals in Figure 6, which shows deviations from a normal distribution, especially on the right tail—suggesting the presence of significant outliers in the predictions.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

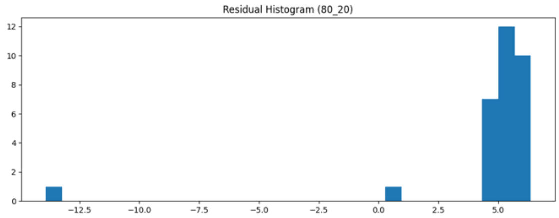

Figure 5. Histogram of Residuals

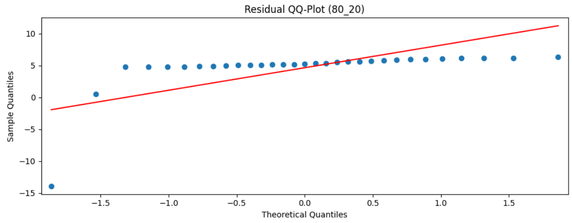

Figure 6. Residuals QQ-Plot

Finally, the scatter plot between predicted and actual values in Figure 7 shows that most of the prediction points tend to fall below the actual values. Some outliers with higher predictions indicate that the model had difficulty forecasting extreme variations in oil and gas production. Overall, although the model learned the general trend in the data, it faced challenges in capturing extreme fluctuations.

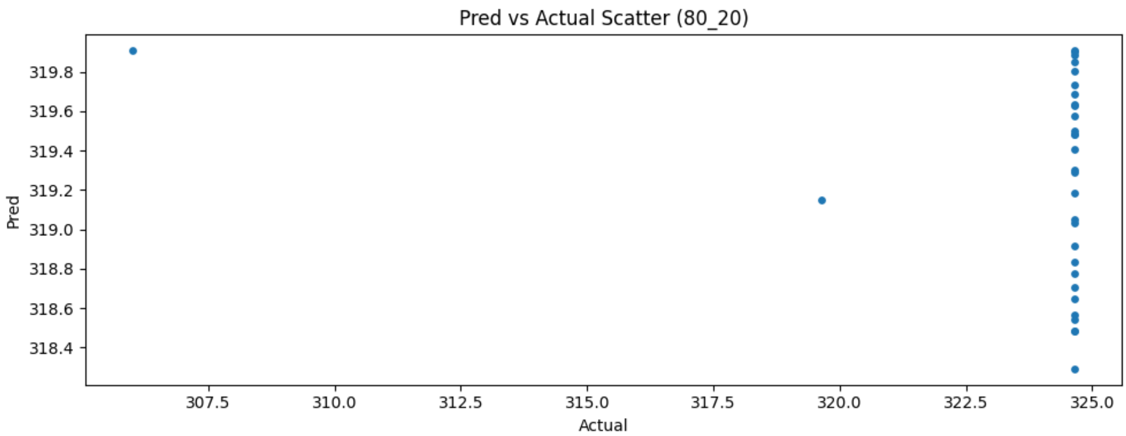

Figure 7. Scatter Plot of Predictions vs Actual Values

2) 90:10 Training Scenario

In the experiment using a 90:10 data split, the model training showed different results compared to the 80:20 scenario. The Loss Curve in Figure 8 reveals a sharp decrease in training loss from the first epoch. However, the validation loss showed more fluctuation compared to the 80:20 model. Although the validation loss eventually stabilized after a few epochs, the initial instability suggests the model faced challenges in optimally understanding the validation data. In the prediction vs actual comparison (Figure 9), the LSTM model was able to follow the general trend of production data over a longer period. However, the predictions were smoother compared to the actual data, which had sharper fluctuations. This indicates that while the model captured overall patterns fairly well, it

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

failed to fully reflect deeper changes in actual values, such as those on 2025-06-15, where the real data remained stable but the predictions deviated slightly.

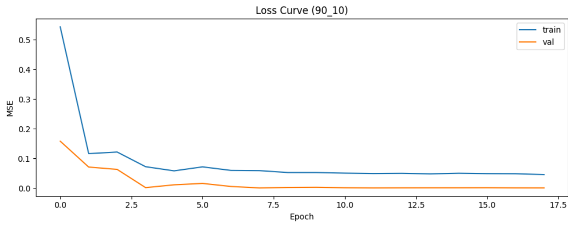

Figure 8. Training and Validation Loss Values

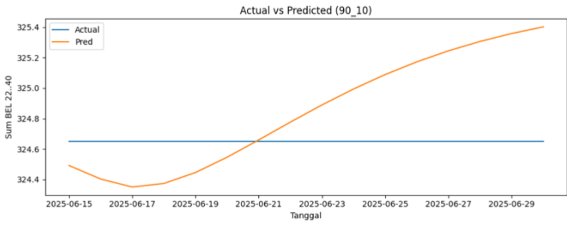

Figure 9. Comparison of Predictions and Actual Data

The Residuals Over Time (Figure 10) displayed smaller fluctuations compared to the 80:20 scenario. Initially, residuals started at small positive values and gradually decreased over time. This pattern indicates that the model was better at predicting long-term trends, although some mismatch remained toward the end of the dataset. The residual histogram (Figure 11) showed a relatively even distribution, with small residuals clustered around zero and a few slightly higher positive values.

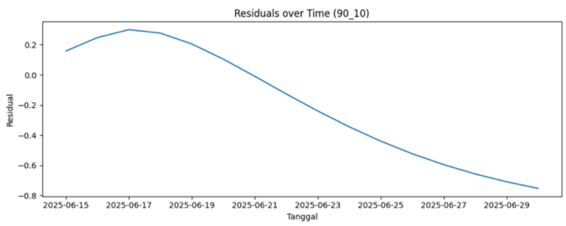

Figure 10. Residual Fluctuations

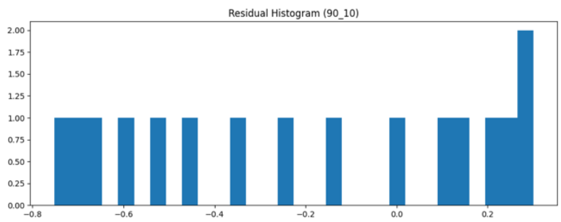

Figure 11. Histogram of Residuals

The QQ-Plot in Figure 12 indicates that most residual points followed a normal distribution, with only minor deviations. This suggests that the model did a better job capturing a consistent residual pattern and produced fewer outliers compared to the 80:20 scenario. Lastly, the scatter plot in Figure 13 shows that most of the predicted values were very close to the actual values, although there were a few larger deviations. This indicates that the model was more successful in making accurate predictions, but still encountered some outliers indicating mispredictions.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

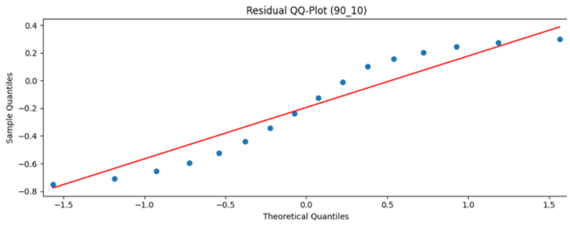

Figure 12. QQ-Plot of Residuals

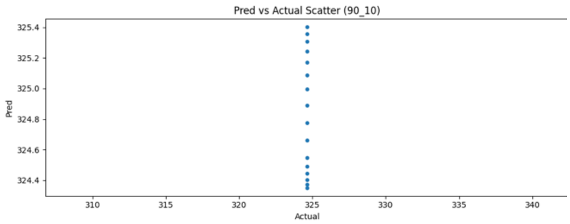

Figure 13. Scatter Plot of Predictions vs Actual Data

Based on the experimental results, the gas prediction model's performance was evaluated using several key metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2). These metrics measure how well the model predicts gas production data. The results are presented in Table 3.

Table 3. Comparison of Training Scenarios

Data Split	RMSE	MAE	MAPE (%)	R^2
80:20	5.831	5.539	1.715	-1.971
90:10	0.420	0.355	0.109	0.000

Based on the metric results in Table 3, the model trained under the 90:10 scenario shows significant improvement compared to the 80:20 scenario. The drastic reductions in RMSE, MAE, and MAPE in the 90:10 split indicate that the model was more capable of accurately predicting oil and gas production, especially in capturing long-term trends. Although the R^2 value in the 90:10 scenario does not indicate a strong ability to explain data variability, the overall decrease in prediction errors suggests that the model is more stable and performs better when trained on a larger dataset. This reinforces the importance of sufficient training data for improving the model's predictive accuracy in real-world production environments.

3) Gas Prediction Model

In this experiment, the LSTM model was applied with both 80:20 (Figure 14) and 90:10 (Figure 15) data splits to forecast gas production at PT Pertamina Hulu Rokan Zone 4. These graphs illustrate a 7-step-ahead forecast, where the historical data is represented by a blue line, and the forecasted values for the next 7-time steps are shown as an orange line. In Figure 14, using the 80:20 scenario, the model successfully identified the general

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

trend in historical gas production data up to March 1, 2025, although significant fluctuations were present in the early periods. For the 7-step-ahead forecast, the model showed stability in its predictions, although it did not fully capture the volatility seen in the historical data.

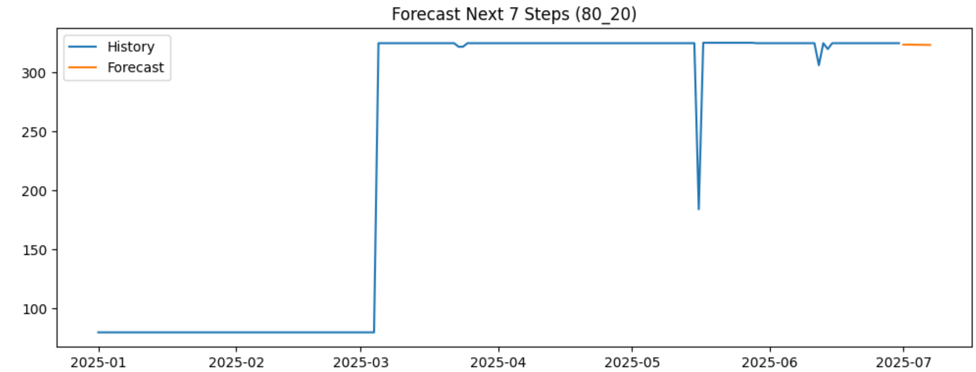

Figure 14. 7-Day Ahead Forecast (80:20)

In Figure 15, based on the 90:10 data split, the prediction results were similar in overall shape and fluctuation to those in Figure 14. The model continued to struggle in forecasting the sharp fluctuations seen at the beginning of the dataset. However, the 7-step-ahead forecasts were smoother and more stable. Based on the prediction results from both the 80:20 and 90:10 scenarios, the LSTM model demonstrated strong performance in forecasting long-term trends in gas production. Nevertheless, the model had difficulty capturing extreme fluctuations in the production data, resulting in more stable forecasts following deep shifts in the data. That said, the 90:10 scenario yielded better results in both stability and accuracy, suggesting that using a larger portion of the data for training allows the model to learn more effectively and capture data patterns more accurately.

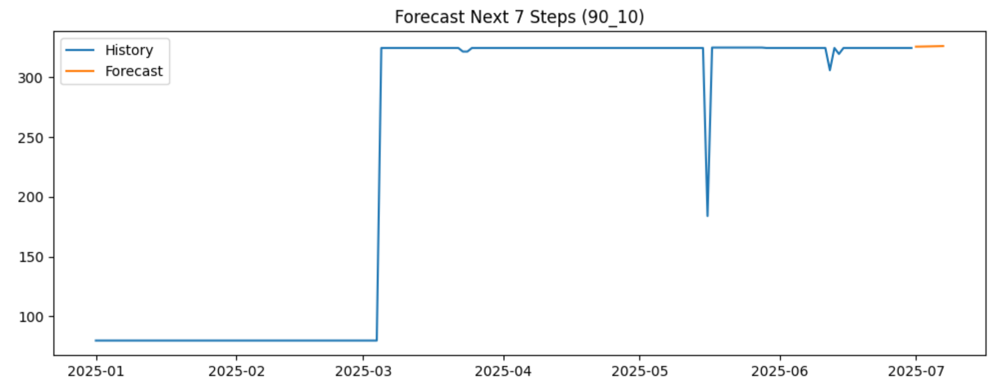

Figure 15. 7-Day Ahead Forecast (90:10)

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.2. Oil Prediction Model Performance

In this experiment, the LSTM model was applied to predict three types of oil production data: Gross, WC (Wellhead Contribution), and Net, using an 80:20 data split. In this scenario, 80% of the data was used for training and 20% for validation. The analysis based on experimental results and several plots is described below for each oil type.

1) Gross Oil Prediction Performance

Figure 16, which shows the training history for gross oil production, indicates that the training loss (blue line) dropped significantly during the first epoch, suggesting the model learned well from the training data. However, the validation loss (orange line) fluctuated more, pointing to overfitting. This fluctuation indicates that while the model learned the training data effectively, it struggled to generalize on unseen data. In Figure 17 (Actual vs Predicted), the model follows the general trend of the data, but the predictions (orange line) are smoother and often fail to capture sharp spikes in the actual data (blue line). The model can predict the direction of trends reasonably well but struggles with sudden shifts, especially early in the data, where large production jumps occurred. Figure 18, showing the residual distribution, reveals that prediction errors varied significantly, with some large residuals indicating inaccurate predictions for specific data points. This confirms that while the model captures general patterns, it still faces challenges in handling extreme changes or spikes in production data. In Figure 19, which shows the most recent actual values and forecasts, the model provides stable future predictions but fails to capture sharp spikes that occurred in early 2025. The predictions follow long-term trends but do not respond well to drastic changes, showing the model's preference for trend-following over volatility detection.

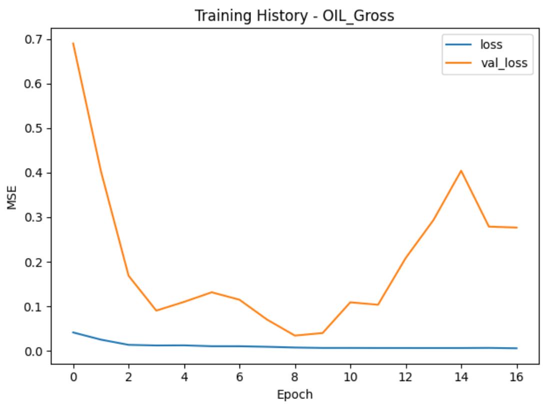

Figure 16. Training History

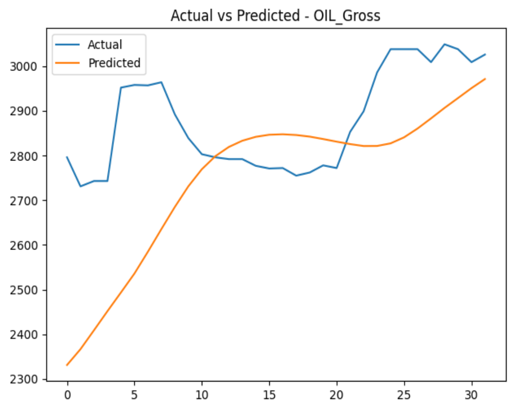

Figure 17. Actual vs Predicted

Vol. 7, No. 4, December 2025

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

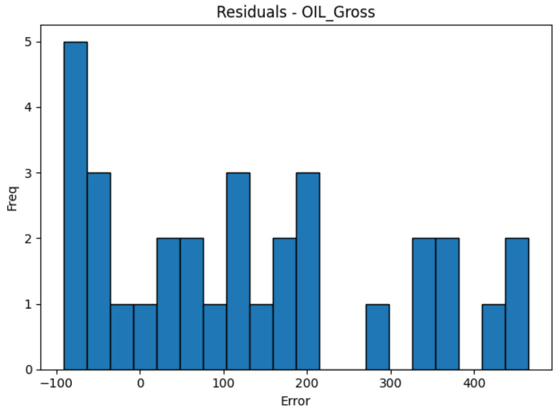

Figure 18. Residual Distribution

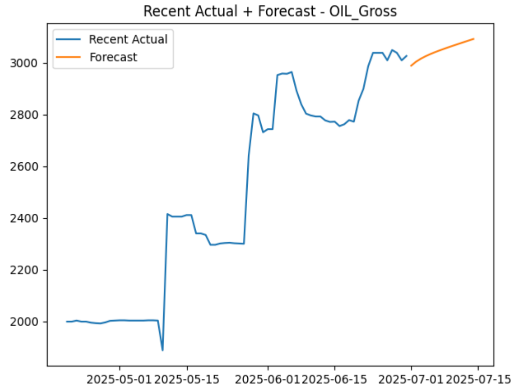

Figure 19. Recent Actual vs Forecast

2) WC Oil Prediction Performance

Figure 20, showing the training history for WC oil production, displays a sharp and steady decline in training loss (blue line). However, similar to the gross model, the validation loss (orange line) fluctuates significantly, again indicating overfitting. This suggests that while the model learns well on the training set, it fails to generalize perfectly on the validation set.

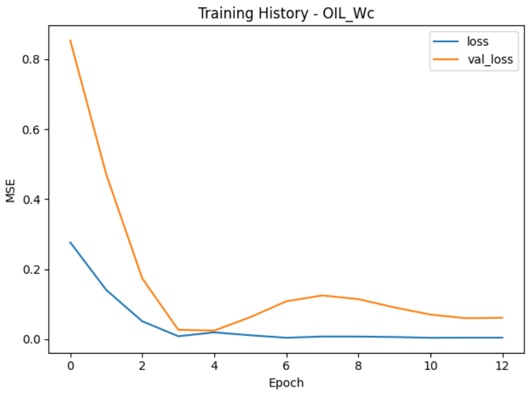

Figure 20. Training History

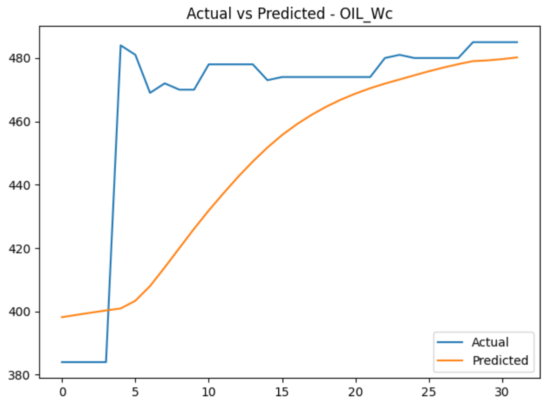

Figure 21. Actual vs Predicted

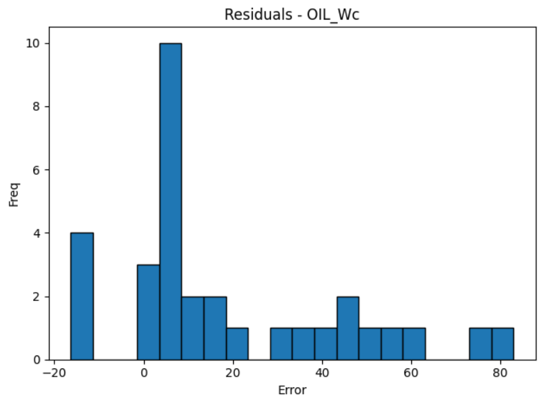

Figure 22. Residual Distribution

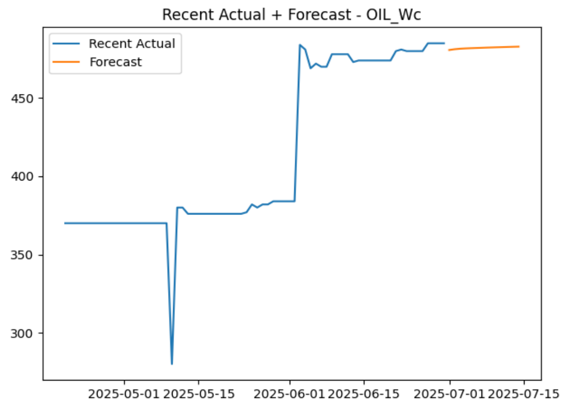

Figure 23. Recent Actual vs Forecast

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

In Figure 21, the comparison of actual vs predicted values for WC oil shows that while the model tracks the general production trend, its predictions (orange line) are not entirely accurate in capturing sharp fluctuations in the actual data (blue line). Predictions are smoother, whereas the real data shows sharper changes, especially around early data spikes. In Figure 22, the residual distribution indicates most errors fall within small to moderate ranges, although there are some outliers with large errors. This suggests the model struggled to forecast points with deep changes or extreme fluctuations, though it accurately predicted the majority of points. Figure 23 presents recent actual data and forecast results. After the large fluctuations seen at the beginning of 2025, the model delivers more stable predictions but continues to struggle in detecting sharp early spikes. This again reflects the model's ability to capture long-term trends, but not drastic changes in WC oil production.

3) Net Oil Prediction Performance

Figure 24, which displays the training history for Net oil production, shows a very sharp decline in training loss with only minor fluctuations. However, like the WC model, validation loss steadily increased, signaling overfitting. The minimal fluctuation in validation loss indicates the model failed to recognize relevant patterns in the validation data.

Figure 24. Training History

Figure 25. Actual vs Predicted

Figure 26. Residual Distribution

Figure 27. Recent Actual vs Forecast

Figure 25, which compares actual vs predicted data for Net oil, shows that the model struggled significantly to predict oil production accurately. Predictions (orange line) could not keep up with changes in the actual data (blue line). After the early data period, model predictions became flat and unresponsive to major rises or falls in actual values. The residual distribution in Figure 26 shows that most errors ranged between 0 and 100, but several large outliers indicate high prediction errors. This shows the model's difficulty in capturing sharp fluctuations in Net oil production, with predictions at times far higher or lower than the actual data. Lastly, Figure 27, which displays the recent actual and forecasted data, reveals significant inaccuracies—especially during large spikes in actual

Vol. 7, No. 4, December 2025

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

production seen after the early data period. The model tends to provide flat, unresponsive forecasts, showing its inability to adapt to abrupt changes in production patterns.

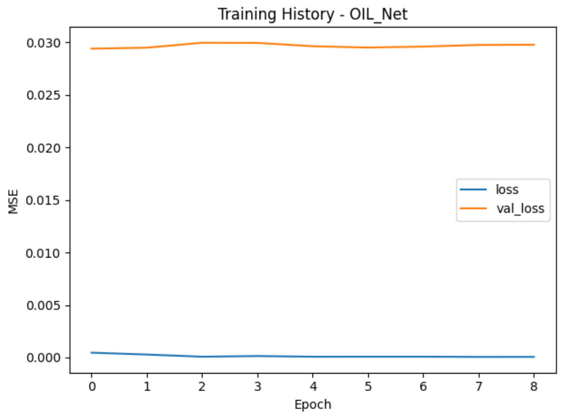

Figure 24. Training History

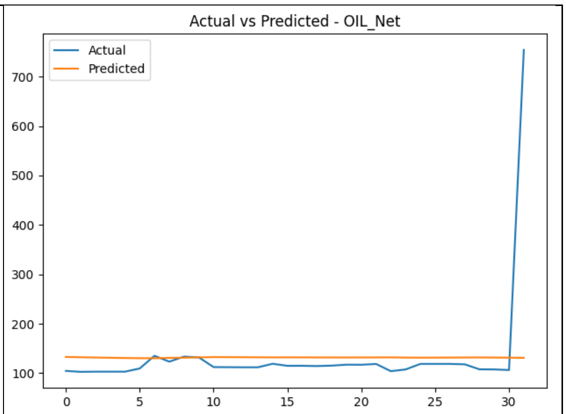

Figure 25. Actual vs Predicted

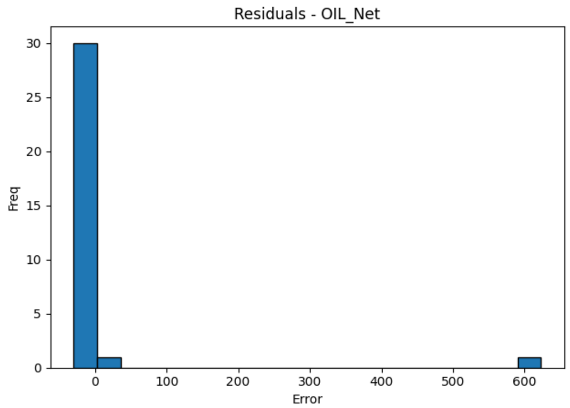

Figure 26. Residual Distribution

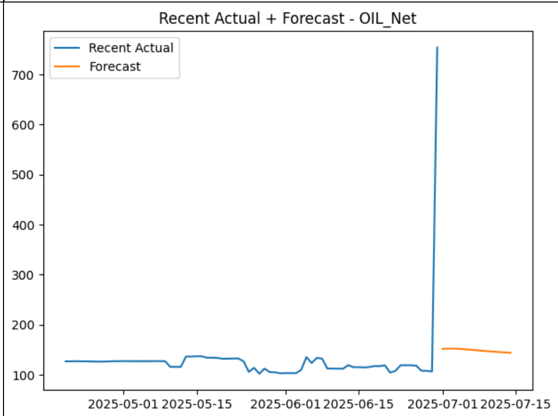

Figure 27. Recent Actual vs Forecast

3.3. Discussion

In this experiment, the Long Short-Term Memory (LSTM) model was applied to predict oil and gas production at PT Pertamina Hulu Rokan Zone 4, using two data split scenarios: 80:20 and 90:10. By implementing an encoder-decoder architecture reinforced with an attention mechanism, the LSTM model was able to capture the sequential patterns in complex production data. However, despite the model performing well in predicting long-term trends, several challenges remained—especially in handling extreme fluctuations in the production data for both oil and gas.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

In the 80:20 scenario, although the model learned well from the training data, signs of overfitting were identified in the validation set. This was evident from the larger fluctuations in the validation loss compared to the training loss, indicating that the model adapted too closely to the training data and struggled to generalize to unseen data. This overfitting made it difficult for the model to accurately capture sharp changes in production, particularly during periods with deep shifts or extreme spikes in oil and gas output. While the model could follow general trends, it struggled at data points with sudden changes, likely caused by external factors not reflected in the historical data.

In the 90:10 scenario, where 90% of the data was used for training, the model demonstrated greater stability and improved prediction accuracy compared to the 80:20 scenario. This improvement was due to the increased volume of training data, which allowed the model to learn more complex and accurate patterns. Nevertheless, even with this improvement, the model still faced difficulties in capturing sharp fluctuations in oil and gas production. Predictions tended to be smoothed, highlighting the model's limitations in responding to sudden spikes or drops in the data.

One of the significant findings from this study is that, despite the model's ability to forecast long-term trends accurately, extreme fluctuations remain a major challenge. In both oil and gas production data, especially at points of abrupt changes, the model was often slow to respond. This limitation may be attributed to the dynamic nature of production data, where sudden increases or drops are often influenced by external factors that are hard to predict using historical data alone. While the LSTM model is effective in learning long-term patterns, it struggles with unforeseen variability, a characteristic feature of oil and gas production data.

The inclusion of the attention mechanism in the LSTM model did improve its focus on the most relevant parts of the data, allowing it to assign greater weight to important sequential patterns. However, despite this enhancement, the attention mechanism was still not sufficient for accurately forecasting sudden, deep changes. In oil and gas production, many external factors—such as market price shifts, regulatory policy changes, or deep operational conditions—can significantly affect production. These are often not captured within the historical dataset used to train the model, limiting its predictive capability.

While the LSTM model with a 90:10 data split showed better performance in recognizing long-term patterns, the experimental results indicate that further strategies are needed to handle extreme fluctuations. One potential solution is to implement regularization techniques to reduce overfitting and enhance the model's ability to generalize. Additionally, the integration of external variables, such as market prices, regulatory updates, or other influencing factors, could help improve prediction accuracy.

The findings from this experiment demonstrate that while LSTM is effective for long-term oil and gas trend forecasting, it continues to face challenges in detecting the frequent extreme fluctuations that characterize production data. The 90:10 scenario yielded better outcomes overall, but the model still struggled with sudden surges or drops. By enhancing training strategies—such as applying regularization methods or incorporating additional features from external data sources—the model's ability to predict extreme variations could be improved. This would result in a more robust and generalized prediction system for dynamic production environments.

4. CONCLUSION

The primary objective of this research was to develop an oil and gas production forecasting model using LSTM with an encoder-decoder architecture, enhanced with an attention mechanism, and to evaluate its performance under two different data split scenarios: 80:20 and 90:10. Based on the experimental results, the LSTM model successfully predicted long-term trends in oil and gas production under both scenarios. The model demonstrated the ability to capture general patterns in the data, although it faced challenges in predicting extreme fluctuations or deep shifts at specific points in the production data. In the 80:20 scenario, the model experienced overfitting on the validation data. This indicated that while the model could learn well from the training set, it struggled to generalize to unseen data. This was evident from the higher fluctuation in validation loss compared to the training loss. However, in the 90:10 scenario, where more training data was available, the model showed significant improvements in stability and prediction accuracy.

Despite the performance improvement in the 90:10 scenario, challenges in capturing extreme fluctuations still persisted, suggesting that the model remained limited in

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

responding to sharp spikes or drops in production. Although LSTM proved effective in capturing long-term trends, it continued to struggle with predicting more drastic changes in oil and gas output. Therefore, while the research objective of developing an accurate and stable forecasting model was achieved, the model still requires further enhancements to effectively address extreme variability. Steps that can be taken to improve prediction accuracy include implementing regularization techniques, optimizing hyperparameter tuning, and integrating relevant external data to capture factors that influence oil and gas production fluctuations.

ACKNOWLEDGMENT

The authors would like to express their deepest gratitude for the research funding provided by Diktisaintek through LLDIKTI Region II and Universitas Bina Darma. This support was invaluable to the smooth execution and success of this research. This study was funded under the Master Contract No.: 123/C3/DT.05.00/PL/2025, Derivative Contract No.: 145/LL2/DT.05.00/PL/2025, and 009/SPK/DRPM/Univ-BD/VI/2025. We greatly appreciate the opportunity to carry out this research, which has undoubtedly contributed positively to the advancement of science and technology.

REFERENCES

[1] J. Elvis, E. Suparman, I. I.-I. L. Review, and undefined 2023, “Hukum Pengelolaan Sumber Daya Minyak Dan Gas Bumi Menurut Pasal 33 Undang-Undang Dasar Tahun 1945,” *Iblam Law Review*, vol. 3, no. 3, pp. 14–29, 2023, doi: 10.52249.

[2] A. Almas, “Analisis Kinerja Algoritma Deep Neural Network dalam Melakukan Prediksi Laju Alir Produksi Minyak pada Lapangan X PT. Pertamina Hulu Rokan,” 2024.

[3] M. Faiz Haikal *et al.*, “Dampak Fluktuasi Pendapatan pada Tingkat Penjualan di Perusahaan BTC Watampone dalam Perspektif Ekonomi Mikro Islam,” *Jurnal Inovasi Ekonomi Syariah dan Akuntansi*, vol. 1, no. 4, pp. 24–33, Jun. 2024, doi: 10.61132/JIESA.V1I4.231.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[4] P. N. Yulisa, M. Al Haris, and P. R. Arum, "Peramalan Nilai Ekspor Migas di Indonesia dengan Model Long Short Term Memory (LSTM) dan Gated Recurrent Unit (GRU)," *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 16, no. 1, pp. 328–341, Jul. 2023, doi: 10.36456/JSTAT.VOL16.NO1.A6121.

[5] K. A. Rijal, A. V. Vitianingsih, Y. Kristyawan, A. L. Maukar, and S. F. A. Wati, "Forecasting Model of Indonesia's Oil & Gas and Non-Oil & Gas Export Value using Var and LSTM Methods," *Jurnal Teknologi dan Manajemen Informatika*, vol. 10, no. 1, pp. 59–69, Jun. 2024, doi: 10.26905/JTMI.V10I1.13127.

[6] A. Arfan and ; Lussiana, "Perbandingan algoritma long short-term memory dengan SVR pada prediksi harga saham di Indonesia," *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*, vol. 13, no. 1, pp. 33–43, 2020, doi: 10.33322/petir.v13i1.858.

[7] A. Alqahtani, M. Ali, X. Xie, and M. W. Jones, "Deep Time-Series Clustering: A Review," *Electronics (Basel)*, vol. 10, no. 23, p. 3001, Dec. 2021, doi: 10.3390/ELECTRONICS10233001.

[8] J. Nurhakiki *et al.*, "Studi Kepustakaan: Pengenalan 4 Algoritma Pada Pembelajaran Deep Learning Beserta Implikasinya," *Pendekar: Jurnal Pendidikan Berkarakter*, vol. 2, no. 1, pp. 270–281, 2024, doi: 10.51903/pendekar.v2i1.598.

[9] B. Penelitian *et al.*, "Penentuan Nilai Terkecil Root Mean Squared Error (RMSE) Metode Holt-Winters Exponential Smoothing Pada Ekspor Kopi Tujuan Jerman," *MathVision: Jurnal Matematika*, vol. 6, no. 1, pp. 37–42, 2024.

[10] I. Nabillah, I. R.-J. of I. System, and undefined 2020, "Mean absolute percentage error untuk evaluasi hasil prediksi komoditas laut," *Journal of Information System*, vol. 5, no. 2, pp. 250–255.

[11] Marie Luthfi Ashari and Mujiono Sadikin, "Prediksi Data Transaksi Penjualan Time Series Menggunakan Regresi Lstm," *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 9, no. 1, pp. 1–10.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[12] L. Wiranda and M. Sadikin, "Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk PT. Metiska Farma," *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, vol. 8, no. 3, pp. 184–196, 2019, doi: 10.23887/JANAPATI.V8I3.19139.

[13] M. Hussein and Y. Azhar, "Prediksi Harga Minyak Dunia Dengan Metode Deep Learning," *Fountain of Informatics Journal*, vol. 6, no. 1, pp. 29–34, 2021, doi: 10.21111/FIJ.V6I1.4446.

[14] Ferdiansyah, S. H. Othman, R. Zahilah Raja Md Radzi, D. Stiawan, Y. Sazaki, and U. Ependi, "A LSTM-Method for Bitcoin Price Prediction: A Case Study Yahoo Finance Stock Market," *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*, pp. 206–210, Oct. 2019, doi: 10.1109/ICECOS47637.2019.8984499.

[15] U. Ependi;, A. F. Rochim, and A. Wibowo, "A Hybrid Sampling Approach for Improving the Classification of Imbalanced Data Using ROS and NCL Methods," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 3, pp. 345–361, 2023, doi: 10.22266/ijies2023.0630.28.

[16] A. Ayub, "Prediksi Kinerja Time Series Well Performance Berdasarkan Long Short-Term Memory (LSTM)," Makassar, 2024.

[17] M. Naufal, "Rendering Dinamis dan Interaksi Pengguna untuk Visualisasi Data Deret Waktu di Realitas X," *Jurnal Teknik ITS*, vol. 13, no. 3, pp. 230–235, 2024.