JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 9, No. 1, June 2026, pp. 290~301
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v9i1.5765

Comparative Forecasting of Kalimantan Palm Oil Production Using Classical, Machine Learning, and Deep Learning Models

I Gusti Agung Putu Mahendra[1], Muhammad Ikhsan Wibowo[2], I Gusti Ayu Nandia Lestari[3]

[1,2] Department of Informatic Engineering, Politeknik Negeri Bengkalis, Riau, Indonesia
[3] Department of Information Technology, Institut Teknologi dan Bisnis STIKOM Bali, Bali, Indonesia

Article Info

Article history:

Received 05 06, 2026
Revised 06 10, 2026
Accepted 06 23, 2026

Keywords:

Palm Oil Production;
Panel time-series Forecasting;
Classical forecasting Models;
Machine Learning;
Deep Learning;

ABSTRACT

Palm oil production forecasting in Kalimantan was important because production patterns differed across regions and changed over time. This study aimed to compare classical statistical, machine learning, and deep learning models for forecasting regional palm oil production using panel time-series data. Annual district and city-level production data from 2010 to 2024 were transformed into a panel format consisting of 56 regional series and 840 observations. Data from 2010 to 2022 were used for training, while data from 2023 to 2024 were used for testing. The evaluated models included Naive Forecast, Autoregressive Integrated Moving Average, Holt Linear Trend, Light Gradient Boosting Machine, Extreme Gradient Boosting, Long Short-Term Memory, Gated Recurrent Unit, and Neural Basis Expansion Analysis for Time Series. The results showed that Long Short-Term Memory achieved the best performance based on root mean squared error of 31,136.64 and coefficient of determination of 0.7695. Meanwhile, Naive Forecast achieved the best mean absolute error, mean absolute percentage error, and symmetric mean absolute percentage error. This study contributed a comparative panel time-series forecasting framework that showed deep learning was effective for reducing large prediction errors, while simple forecasting remained competitive for stable regional patterns.

Corresponding Author:

I Gusti Agung Putu Mahendra
Department of Informatic Engineering
Politeknik Negeri Bengkalis
Riau, Indonesia
Email: agungmahendra@polbeng.ac.id

1. Introduction

Oil palm is one of the strategic plantation commodities that plays an important role in supporting regional economies, processing industries, and agricultural sector development planning in Indonesia[1]. In Kalimantan, oil palm production is distributed across various provinces and regencies/municipalities with different production characteristics[2]. Differences in regional conditions, plantation area, processing capacity, and annual production dynamics cause oil palm production patterns to be not always linear and stable[3]. Therefore, oil palm production forecasting is important to support production planning, decision-making, policy evaluation, and more measurable plantation commodity supply chain management.

The Kalimantan oil palm production data used in this study covered 56 regency/municipality areas during the 2010–2024 period, forming panel time-series data with 840 observations. The data were then divided into training data from 2010–2022 and testing data from 2023–2024 to evaluate the model's ability to predict production in periods that were not used during the training process. This data structure not only represents changes in production over time but also reflects production variation across regions, making a panel time-series forecasting approach relevant to apply[4], [5].

The main problem in forecasting oil palm production is the high variation in production patterns across regions and sharp trend changes in certain periods[6]. Based on historical data, total production in Kalimantan showed significant changes, particularly after 2017. Total production, which previously reached approximately 12.08 million tons in 2017, sharply declined to around 1.52 million tons in 2018, and then remained within the range of approximately 2.39–2.78 million tons during 2021–2024. This condition indicates the possibility of changes in production patterns, changes in recording coverage, or structural breaks that may affect the stability of forecasting models. Therefore, using only one type of model may not be sufficient to comprehensively describe predictive performance[7].

Various methods have been used in time-series forecasting studies. Classical models such as Naive Forecast, Holt, and ARIMA are commonly used as baselines because they are simple and easy to interpret[8]. ARIMA and exponential smoothing are widely discussed in classical forecasting literature for modeling historical patterns and trends in time-series data[9]. However, classical methods generally have limitations in capturing nonlinear relationships and heterogeneity across regions. Along with the development of machine learning, gradient boosting-based models such as XGBoost and LightGBM have been widely used for tabular data because they are able to effectively learn nonlinear patterns and feature interactions. XGBoost was developed as a scalable tree boosting system and has been widely used to achieve high predictive performance in various data mining tasks, while LightGBM was designed as a more efficient implementation of gradient boosting decision trees through techniques such as Gradient-based One-Side Sampling and Exclusive Feature Bundling[10].

In addition to machine learning, deep learning approaches have also been increasingly used for time-series forecasting because of their ability to learn complex temporal patterns. LSTM was developed to overcome the vanishing gradient problem in recurrent neural networks and was designed to retain information over longer time spans[11]. GRU is a recurrent neural network variant with a simpler gating mechanism and has been reported in several studies to achieve performance comparable to LSTM in sequence modeling. Meanwhile, N-BEATS is a deep learning architecture for forecasting that uses backward and forward residual links as well as fully connected layers, and is designed to be applicable across various time-series domains[12].

Although various forecasting models are available, previous studies on oil palm production data in Kalimantan have tended to use classical approaches such as Naive Forecast, Holt, and ARIMA. These approaches are important as baselines, but they are not sufficient to assess whether machine learning and deep learning models can improve performance on panel data of oil palm production. In addition, few studies have systematically compared classical, machine learning, and deep learning models on oil palm production data across regencies/municipalities in Kalimantan. This gap is important because oil palm production data has panel characteristics, containing both time and regional dimensions simultaneously.

More specifically, the research gap in the context of oil palm production forecasting in Kalimantan can be identified in three aspects. First, most previous approaches have focused on aggregate or single time-series forecasting, while limited attention has been given to panel time-series modeling that simultaneously captures temporal changes and regional heterogeneity across regencies/municipalities. Second, prior studies have not sufficiently compared different model families under the same experimental setting, particularly classical statistical models, machine learning models, and deep learning models. Third, regional variability in oil palm production has not been deeply analyzed, even though each province and regency/municipality in Kalimantan may have different production scales, trend stability, and error characteristics. Therefore, a comparative panel time-series forecasting approach is needed to evaluate which model is more suitable for heterogeneous regional oil palm production data in Kalimantan.

Based on these problems, this study proposes a comparative approach to evaluate the performance of several groups of forecasting models, namely classical models, machine learning models, and deep learning models. The models used include Naive Forecast, ARIMA, Holt, LightGBM, XGBoost, LSTM, GRU, and N-BEATS. Evaluation was conducted using MAE, RMSE, MAPE, sMAPE, and R^2 metrics so that model performance could be assessed from various error perspectives[13]. The novelty of this study lies in the application of a comparative panel time-series forecasting approach to oil palm production in Kalimantan by comparing three groups of models simultaneously, namely classical statistical models, machine learning models, and deep learning models. Unlike previous studies that focused only on classical forecasting models, this study utilizes the panel structure of cross-regional data and evaluates modern models such as LightGBM, XGBoost, LSTM, GRU, and N-BEATS. Thus, this study not only produces predictive models but also provides an understanding of which models are most suitable for the characteristics of Kalimantan oil palm production data based on various evaluation metrics.

In general, this study is expected to provide both methodological and practical contributions. From a methodological perspective, this study presents a comprehensive comparison between classical, machine learning, and deep learning models on panel time-series data of oil palm production. From a practical perspective, the forecasting results can serve as an initial basis for local governments, plantation sector managers, and related stakeholders in understanding oil palm production trends, supporting regional agricultural planning, preparing supply chain strategies, and strengthening data-driven decision-making in the palm oil sector.

2. Research Method

This study used a quantitative experimental research design with a comparative forecasting approach. The objective was to evaluate the performance of classical, machine learning, and deep learning models for forecasting palm oil production in Kalimantan using panel time-series data.

The forecasting task was formulated as a supervised time-series prediction problem. The input data consisted of regional annual palm oil production records, while the output was the predicted production value for future years. The models were evaluated using the same training and testing scheme to ensure a fair comparison. The dataset contained 56 regional time series, covering the period 2010-2024, with a total of 840 observations after transformation into panel time-series format. The data were divided into training data from 2010-2022 and testing data from 2023-2024, with a forecast horizon of two years.

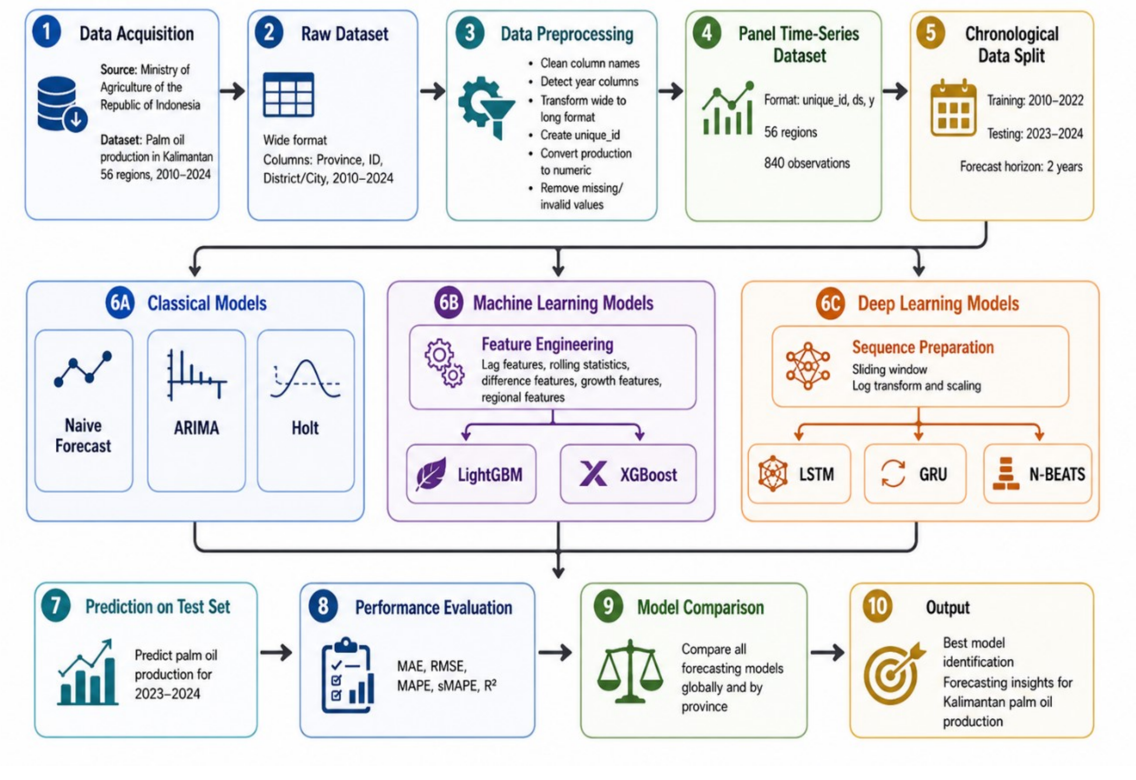

Figure 1. Research Flow

2.1 Data Acquisition

The data used in this study were obtained from the official database of the Ministry of Agriculture of the Republic of Indonesia. The dataset contains annual palm oil production data from districts/cities in Kalimantan as summerized Table 1. The original dataset was arranged in wide format, where each year from 2010 to 2024 was represented as a separate column. The original attributes in the dataset were:

Table 1. Dataset Acquisition

Attribute	Description
Provinsi	Province name in Kalimantan
ID	Regional administrative code
Kabupaten_Kota	District/city name
Ton_Ha	Measurement unit
2010–2024	Annual palm oil production in tons

The raw data had 56 rows and 19 columns, consisting of regional identity columns and annual production columns from 2010 to 2024.

2.2 Data Preprocessing

The preprocessing stage was conducted to transform the raw dataset into a structured panel time-series format suitable for comparative forecasting experiments. The original dataset was arranged in wide format, where each year from 2010 to 2024 was represented as a separate column. The year columns were first identified automatically and then transformed into long format using a wide-to-long transformation process. In the transformed dataset, each observation represented one region in one year.

A regional identifier, named unique_id, was created by combining the province and district/city names. The year variable was converted into ds, while annual palm oil production was converted into the target variable y. Production values were converted into numeric format to ensure compatibility with statistical, machine learning, and deep learning models. Missing values and invalid production values were checked before modeling. Records with missing or invalid target values were removed to avoid bias in model training and evaluation.

For machine learning models, additional lag-based features were generated to represent historical production patterns. These features included lag values from the previous one to five years, rolling mean, rolling standard deviation, production differences, growth rates, and regional identifiers. These engineered features were used by LightGBM and XGBoost to learn nonlinear relationships between historical production behavior and future production values. For deep learning models, the dataset was prepared using a sliding-window approach, where the previous five years of production data were used as input to predict the next two years. This preprocessing design ensured that all models were trained and evaluated using the same chronological data structure [14].

Table 2. Dataset Transform

Column	Description
unique_id	Regional time-series identifier
ds	Year
y	Palm oil production in tons

2.3 Data Splitting

The dataset was divided chronologically to avoid data leakage. The training set used observations from 2010 to 2022, while the testing set used observations from 2023 to 2024.

Table 3. Split Dataset

Data Partition	Period	Number of Observations	Purpose
Training data	2010–2022	728	Model training
Testing data	2023–2024	112	Model evaluation
Forecast horizon	2 years	-	Predicting 2023 and 2024

The chronological split was selected because forecasting problems must preserve temporal order. Therefore, future values were not used during model training.

2.4 Model Configuration

To improve the reproducibility of the experiment, the main configuration of each model is described as follows. The Naive Forecast model used the last observed value in the training period as the prediction for the testing period. ARIMA was implemented separately for each regional time series using an order of (1,1,1). Holt Linear Trend was also applied separately to each region to capture level and trend components.

For machine learning models, LightGBM and XGBoost were trained using lag-based and rolling statistical features. The feature set included lag-1 to lag-5, rolling mean, rolling standard deviation, difference features, growth features, year index, province, district/city, and regional identifier. Recursive forecasting was used to predict the 2023–2024 testing period, where the prediction for 2023 was used as input for forecasting 2024.

For deep learning models, LSTM and GRU were trained using a sliding-window input of five years and a forecast horizon of two years. The production values were transformed using logarithmic transformation and standard scaling before training to reduce the effect of large value differences among regions. Both models used one recurrent layer, 64 hidden units, a batch size of 32, and 300 training epochs. N-BEATS was trained as a global panel forecasting model using a forecast horizon of two years, an input size of five years, and a lightweight architecture consisting of trend and identity stacks. The N-BEATS model was trained for 500 maximum steps with a learning rate of 0.001.

2.4.1 Naive Forecast

The Naive Forecast model was used as the baseline. This model predicts future production using the last observed value from the training period[15]. where y_t is the last observed production value in the training data.

$$\widehat{y_{t+h}} = y_t \tag{1}$$

2.4.2 ARIMA

ARIMA was applied separately to each regional time series. The ARIMA model is suitable for classical univariate time-series forecasting because it models autoregressive, differencing, and moving average components[16]. The ARIMA model can be represented as:

$$ARIMA(p, d, q) \tag{2}$$

Parameter	Description
p	: Autoregressive order
d	: Differencing order
q	: Moving average order

In this study, ARIMA was trained for each region using the historical data from 2010-2022, and then used to forecast 2023-2024.

2.4.3 Holt Linear Trend

Holt Linear Trend was used to model trend-based time-series patterns. This method estimates both the level and trend components of the data. Holt forecasting was also applied separately for each region [17]. The general Holt formulation is:

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \tag{3}$$
$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \tag{4}$$
$$\widehat{y_{t+h}} = l_t + hb_t \tag{5}$$

where l_t is the level, b_t is the trend, and hhh is the forecast horizon.

2.4.4 LightGBM and XGBoost

LightGBM and XGBoost were used as machine learning models. Unlike ARIMA and Holt, these models were trained as global panel models using engineered features from all regions. The features used for LightGBM and XGBoost included:

Table 4. Paramater LightGBM and XGBoost

Feature Type	Features
Regional features	Province, district/city, unique_id
Time features	Year, year index
Lag features	lag-1, lag-2, lag-3, lag-4, lag-5
Rolling features	rolling mean-3, rolling mean-5, rolling standard deviation-3, rolling standard deviation-5
Difference features	diff-1, diff-2
Growth features	growth-1, growth-2
Statistical features	minimum, maximum, and median of the last five years

The machine learning training data contained lag-based and rolling statistical features for each regional time series. For multi-step prediction, LightGBM and XGBoost used a recursive forecasting strategy. The prediction for 2023 was used as input to forecast 2024, preventing the model from using actual future values during testing [18].

2.4.5 LSTM

Long Short-Term Memory was used as a deep learning model to capture sequential dependencies in the production data. LSTM was trained using sliding windows from the training data[19]. The input sequence length was set to five years, and the output horizon was two years:

$$X = [y_{t-4}, y_{t-3}, y_{t-2}, y_{t-1}, y_t] \tag{6}$$
$$Y = [y_{t+1}, y_{t+2}] \tag{7}$$

Before training, the target values were transformed using logarithmic transformation and standard scaling to stabilize the range of production values.

2.4.6 GRU

Gated Recurrent Unit was used as another recurrent deep learning model. GRU has a simpler gating mechanism than LSTM but can still capture temporal dependencies in sequential data[20]. The GRU model used the same input-output configuration as LSTM:

Table 5. Component GRU

Component	**Value**
Input window	5 years
Forecast horizon	2 years
Training target	2023–2024 prediction
Data transformation	log1p and standard scaling

2.4.7 N-BEATS

N-BEATS was used as a deep learning forecasting model based on fully connected neural networks and residual learning. In this study, N-BEATS was trained as a global panel time-series model using all regional series simultaneously[21].

The N-BEATS model was configured with a two-year forecast horizon for evaluation. The model was trained using the panel time-series format consisting of unique_id, ds, and y. The N-BEATS training process successfully generated predictions for the 2023–2024 testing period.

3. Result and Discussion

This section presents the experimental results and discussion of the comparative panel time-series forecasting models for palm oil production in Kalimantan. The results are presented through descriptive statistics, model comparison tables, provincial-level analysis, and forecasting interpretation. The evaluated models include Naive Forecast, ARIMA, Holt, LightGBM, XGBoost, LSTM, GRU, and N-BEATS.

3.1 Dataset Transformation and Experimental Setting

The original dataset consisted of palm oil production records from 56 districts/cities in Kalimantan for the period 2010-2024. The raw data were initially arranged in wide format, where each year was represented as a separate column. The dataset was then transformed into a panel time-series format consisting of three main variables: unique_id, ds, and y. The unique_id represents each district/city, ds represents the year, and y represents palm oil production in tons.

After preprocessing, the final dataset contained 840 observations, representing 56 regional time series with 15 annual observations for each region. The dataset was split chronologically, where data from 2010-2022 were used as training data and data from 2023-2024 were used as testing data. This splitting strategy was applied to preserve the temporal order of the forecasting task and to avoid data leakage.

Table 6. Dataset and experimental setting

Component	Description
Number of regions	56 districts/cities
Period	2010–2024
Total observations	840
Training period	2010–2022
Testing period	2023–2024
Forecast horizon	2 years
Target variable	Palm oil production in tons

The use of panel time-series data is important because palm oil production in Kalimantan is not only influenced by temporal changes but also by regional differences. Each district/city may have different production patterns, production scale, and trend characteristics. Therefore, the forecasting problem requires models that can capture both temporal patterns and regional heterogeneity.

3.2 Global Model Performance

The performance of all forecasting models was evaluated using five metrics: MAE, RMSE, MAPE, sMAPE, and R^2. Table 7 presents the global comparison results across all regions and testing years.

Table 7. Global performance comparison of forecasting models

Rank by RMSE	Model	MAE	RMSE	MAPE (%)	sMAPE (%)	R^2

1	LSTM	17,381.81	31,136.64	45.44	68.62	0.7695
2	GRU	17,914.31	34,293.47	43.32	65.29	0.7204
3	N-BEATS	14,145.98	36,042.39	29.65	47.30	0.6911
4	LightGBM	20,903.86	39,349.90	70.71	71.70	0.6318
5	Naive Forecast	11,674.02	43,249.75	24.05	22.43	0.5552
6	ARIMA	16,559.22	45,286.35	119.59	48.66	0.5124
7	XGBoost	24,481.89	48,944.31	55.43	72.27	0.4304
8	Holt	25,475.90	57,728.38	80.93	77.09	0.2076

The results show that LSTM achieved the best performance based on RMSE and R^2, with an RMSE of 31,136.64 and an R^2 of 0.7695. This indicates that LSTM was more effective in reducing large prediction errors compared with the other models. GRU ranked second with an RMSE of 34,293.47, followed by N-BEATS with an RMSE of 36,042.39. These findings indicate that deep learning models generally performed better than classical and machine learning models when evaluated using RMSE and R^2.

However, the results also show that Naive Forecast produced the lowest MAE, MAPE, and sMAPE. Naive Forecast obtained an MAE of 11,674.02, MAPE of 24.05%, and sMAPE of 22.43%. This indicates that, although LSTM reduced large errors more effectively, the simple Naive Forecast model was still competitive for many regions with relatively stable production patterns. Therefore, the best model depends on the evaluation metric used. Figure 2. presents the global RMSE comparison of all forecasting models used in this study. The results show that LSTM achieved the lowest RMSE, followed by GRU and N-BEATS. This indicates that deep learning models generally provided better performance in reducing large forecasting errors compared with classical and machine learning models. Holt produced the highest RMSE, indicating weaker forecasting performance in the testing period. Although Naive Forecast did not achieve the lowest RMSE, it still remained competitive compared with ARIMA, XGBoost, and Holt.

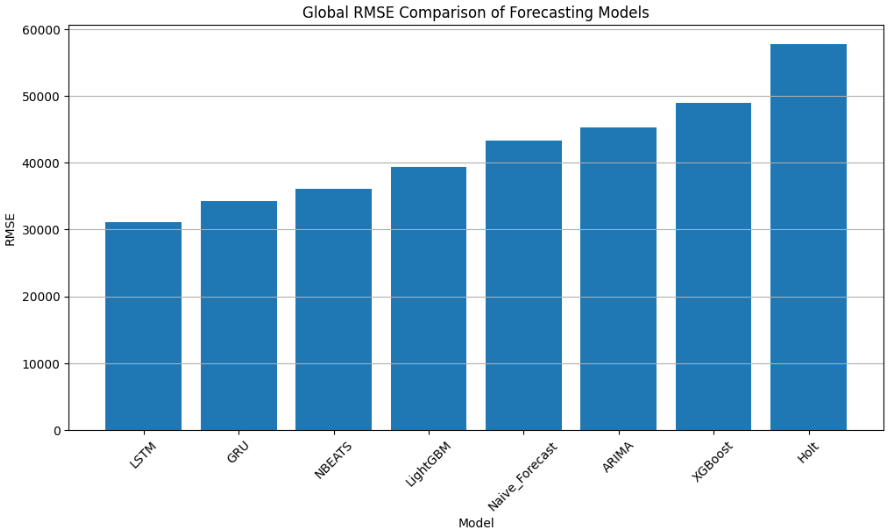

Figure 2. Presents the global RMSE comparison of all forecasting

3.3 Discussion of Global Model Performance

The superior RMSE and R^2 values of LSTM suggest that the model was able to capture sequential patterns in the panel time-series data more effectively than other models. LSTM is designed to learn temporal dependencies through memory cells and gating mechanisms, which makes it suitable for time-series forecasting tasks [22]. In this study, the model benefited from the sliding-window approach, where the previous five years of production data were used to predict production in the next two years.

GRU also showed strong performance, ranking second based on RMSE. This result is reasonable because GRU has a gating mechanism similar to LSTM but with a simpler architecture. The relatively close performance between LSTM and GRU indicates that recurrent neural network models are suitable for modeling temporal dependencies in palm oil production data.

N-BEATS ranked third based on RMSE and R^2. Although N-BEATS did not outperform LSTM and GRU, its performance was still competitive. N-BEATS achieved lower RMSE than LightGBM, Naive Forecast, ARIMA, XGBoost, and Holt. This indicates that N-BEATS can be considered a promising deep learning model for panel time-series forecasting, although its performance may depend on data length, model configuration, and the stability of production patterns.

In contrast, the classical models ARIMA and Holt produced lower performance than most deep learning models. ARIMA obtained an RMSE of 45,286.35, while Holt obtained the highest RMSE of 57,728.38. These results suggest that classical univariate time-series models may have difficulty capturing regional heterogeneity and irregular changes in production patterns across multiple districts/cities.

The machine learning models, LightGBM and XGBoost, did not outperform the deep learning models in the global evaluation. LightGBM performed better than Naive Forecast in terms of RMSE but produced higher MAE, MAPE, and sMAPE. XGBoost performed lower than LightGBM in the global comparison. This result may be influenced by the limited number of annual observations per region, where each region only contains 15 data points. Figure 3 illustrates the comparison between actual and predicted total palm oil production in Kalimantan for the testing period 2023–2024. The actual production increased from 2023 to 2024. Several models, such as Naive Forecast, ARIMA, Holt, and N-BEATS, tended to produce higher total predictions than the actual values. In contrast, LightGBM, XGBoost, LSTM, and GRU tended to underestimate total production. This result indicates that each model captured different production patterns, and the prediction behavior varied depending on the model architecture and learning mechanism.

The evaluation results show different model rankings depending on the metric used. LSTM achieved the best performance based on RMSE and R^2, while Naive Forecast achieved the best performance based on MAE, MAPE, and sMAPE. This difference does not indicate inconsistency in the experiment, but rather reflects the different sensitivity of each evaluation metric.

RMSE gives greater penalty to large prediction errors because the errors are squared before averaging. Therefore, the lower RMSE obtained by LSTM indicates that this model was more effective in reducing large forecasting deviations, especially in regions with high production values or more dynamic production patterns. The R^2 value of 0.7695 also indicates that LSTM explained a larger proportion of production variability than the other models.

On the other hand, MAE, MAPE, and sMAPE evaluate forecasting errors from different perspectives. MAE measures the average absolute error, while MAPE and sMAPE measure relative percentage errors. The superior performance of Naive Forecast on these metrics indicates that many regional production series were relatively stable during the testing period. In such cases, the last observed value from the training period can be a strong predictor for the next year. Therefore, Naive Forecast remained competitive for stable regions, even though it was less effective in reducing large errors at the global level.

These findings show that model selection should depend on the forecasting objective. If the main objective is to minimize large errors in high-production regions, LSTM is more appropriate. However, if the objective is to obtain stable and simple forecasts for regions with relatively constant production patterns, Naive Forecast can still be used as a strong baseline.

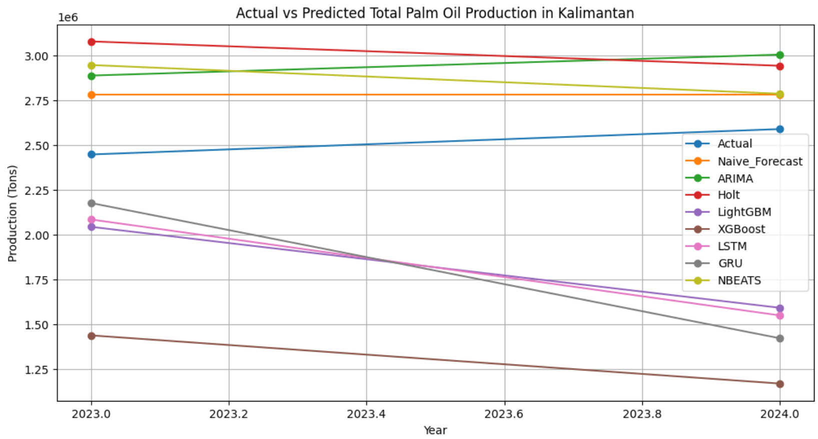

Figure 3. Comparison between actual and predicted

3.4 Provincial-Level Model Performance

In addition to global evaluation, the models were also evaluated by province to identify regional differences in forecasting performance. Table 8 summarizes the best-performing model for each province based on RMSE.

Table 8. Best model by province based on RMSE

Province	Best Model	RMSE	R^2
Kalimantan Barat	Naive Forecast	12,738.12	0.9599
Kalimantan Selatan	Naive Forecast	3,710.66	0.9912
Kalimantan Tengah	Naive Forecast	11,863.32	0.9825

Kalimantan Timur	XGBoost	19,696.65	0.7749
Kalimantan Utara	Naive Forecast	1,446.52	0.9968

The provincial evaluation shows that the best model varied across regions. Naive Forecast achieved the best RMSE in Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, and Kalimantan Utara. Meanwhile, XGBoost achieved the best RMSE in Kalimantan Timur. This indicates that no single model was consistently superior across all provinces.

The strong performance of Naive Forecast in several provinces suggests that the production values in those provinces were relatively stable during the testing period. When the production pattern is stable, the last observed value from the training period can serve as a strong predictor for the next period. This explains why Naive Forecast achieved low error in Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, and Kalimantan Utara.

Kalimantan Timur showed a different pattern. In this province, XGBoost obtained the lowest RMSE of 19,696.65, while Naive Forecast produced a much higher RMSE of 100,159.31. This indicates that Kalimantan Timur may have more dynamic production patterns, making a feature-based machine learning model more suitable than a simple last-value baseline. XGBoost was able to use lag features, rolling statistics, difference features, growth features, and regional features to capture patterns that were not adequately represented by Naive Forecast.

3.5 Interpretation of Deep Learning Models

The global results indicate that deep learning models were generally strong in reducing large forecasting errors. LSTM, GRU, and N-BEATS occupied the top three positions based on RMSE. This suggests that deep learning models are useful for capturing complex temporal relationships in panel time-series data.

LSTM achieved the lowest RMSE, showing its ability to model temporal dependencies across regional production patterns. GRU followed closely behind LSTM, indicating that a simpler recurrent architecture can still produce competitive performance. N-BEATS also provided a strong result, although it did not outperform LSTM and GRU[23], [24].

However, deep learning models did not dominate all evaluation metrics. The Naive Forecast model still achieved the best MAE, MAPE, and sMAPE. This means that while deep learning models are better at reducing large errors, they may still produce relatively high percentage errors in regions with low or stable production values. Therefore, the results should not be interpreted as evidence that deep learning is always superior. Instead, the findings show that deep learning models are advantageous for minimizing large-scale errors, while simple models may remain effective for stable regional series.

3.6 Interpretation of Classical and Machine Learning Models

The classical models showed mixed results. ARIMA performed better than Holt but was still weaker than the leading deep learning models based on RMSE. The ARIMA model also produced a high MAPE value, indicating that its percentage error was unstable for some regions. Holt produced the weakest global performance, with the highest RMSE and lowest R^2.

The machine learning models also showed varied performance. LightGBM ranked fourth based on RMSE, outperforming Naive Forecast, ARIMA, XGBoost, and Holt in terms of RMSE. However, LightGBM produced relatively high MAPE and sMAPE values. XGBoost ranked lower globally but performed best in Kalimantan Timur. This suggests that machine learning models may be useful for specific regional patterns, especially when production behavior is less stable and requires additional engineered features.

The difference between global and provincial results highlights the importance of evaluating forecasting models at multiple levels. A model that performs well globally may not be the best model for every province. Therefore, regional-level analysis is necessary to obtain more practical insights for decision-making [25].

3.7 Forecasting Implications

The experimental results have several implications for palm oil production forecasting in Kalimantan. First, LSTM is the most suitable model when the main objective is to reduce large prediction errors, as shown by its lowest RMSE and highest R^2. Second, Naive Forecast remains important as a strong baseline, especially for provinces with stable production patterns. Third, XGBoost may be more suitable for regions with dynamic production behavior, as shown in the case of Kalimantan Timur.

Additional forecasting results for 2025–2027 using the trained forecasting framework showed that total palm oil production in Kalimantan was predicted to remain relatively stable. The total predicted production was 2,665,795.75 tons in 2025, 2,665,549.00 tons in 2026, and 2,698,897.00 tons in 2027. These results suggest a slight increase in total production by 2027.

Table 9. Total forecasted palm oil production in Kalimantan

Year	Predicted Production (tons)
2025	2,665,795.75
2026	2,665,549.00
2027	2,698,897.00

The forecast results indicate that palm oil production in Kalimantan is expected to remain relatively stable in the short term. However, the prediction should be interpreted carefully because the historical data show a major shift in production patterns after 2017. The total production increased from 3.87 million tons in 2010 to 12.08 million tons in 2017, then declined sharply to 1.52 million tons in 2018 before stabilizing around 2.39-2.78 million tons in 2021-2024. This structural change may affect model stability and forecasting accuracy.

3.8 Overall Discussion

Overall, the results show that model performance depends strongly on the evaluation metric and the regional characteristics of the data. Based on RMSE and R^2, LSTM was the best global model, indicating that it was effective in reducing large forecasting errors. However, based on MAE, MAPE, and sMAPE, Naive Forecast was the best model, showing that simple forecasting methods can remain competitive when the data are stable.

The finding that Naive Forecast performed strongly in several provinces is important. It suggests that some regions have production patterns where the latest historical value is already a strong predictor. In such cases, more complex models may not always provide better performance. On the other hand, the strong performance of XGBoost in Kalimantan Timur indicates that machine learning models can be useful in regions with more variable production patterns.

The results also demonstrate the value of a comparative forecasting approach. If the study only evaluated one model, such as N-BEATS or LSTM, the conclusion could be misleading. By comparing classical, machine learning, and deep learning models, this research provides a more comprehensive understanding of which model is suitable under different evaluation criteria and regional conditions.

3.9 Practical and Policy Implications

The findings of this study provide several practical and policy implications for palm oil production planning in Kalimantan. First, the forecasting results can support regional agricultural planning by providing estimated production trends for districts/cities and provinces. These forecasts can help local governments and agricultural agencies anticipate production changes and prepare appropriate policy responses.

Second, the comparison of forecasting models can assist decision-makers in selecting appropriate forecasting methods based on regional characteristics. For regions with relatively stable production patterns, simple models such as Naive Forecast may be sufficient and easy to implement. However, for regions with more dynamic production behavior, deep learning or machine learning models may be more suitable because they can better capture complex temporal and regional patterns.

Third, the forecasting results can support supply chain preparation in the palm oil sector. More accurate production forecasts can help stakeholders plan processing capacity, logistics, distribution, and market supply. This is important because palm oil production affects not only agricultural planning but also industrial processing and regional economic activity.

Fourth, the provincial-level evaluation provides useful insight for regional policy-making. The results show that no single model dominated all provinces. Therefore, forecasting systems should not rely only on one global model but should consider regional characteristics and model performance at the provincial or district/city level. This approach can improve the relevance of forecasting results for local decision-making. Finally, the study provides a methodological basis for developing data-driven monitoring systems for palm oil production. Such systems can be used by government institutions, plantation agencies, and industry stakeholders to monitor production trends, identify potential production instability, and support evidence-based agricultural policy.

4. Conclusion

This study successfully addressed the research objective stated in the Introduction, namely to compare classical statistical, machine learning, and deep learning models for forecasting palm oil production in Kalimantan using regional panel time-series data. The dataset consisted of 56 regional series from 2010 to

2024, with a total of 840 observations. Data from 2010–2022 were used for training, while data from 2023–2024 were used for testing.

The results showed that model performance varied depending on the evaluation metric. Based on RMSE and R^2, LSTM achieved the best global performance, with an RMSE of 31,136.64 and an R^2 of 0.7695. This indicates that LSTM was more effective in reducing large forecasting errors. However, Naive Forecast achieved the best performance based on MAE, MAPE, and sMAPE, with an MAE of 11,674.02, MAPE of 24.05%, and sMAPE of 22.43%. This finding shows that simple forecasting methods can remain competitive for regions with relatively stable production patterns.

The provincial-level analysis also showed that no single model dominated all regions. Naive Forecast performed best in Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, and Kalimantan Utara, while XGBoost achieved the best performance in Kalimantan Timur. This result confirms that regional production characteristics strongly influence forecasting performance, so model selection should consider both evaluation metrics and regional behavior.

The forecasting results for 2025–2027 indicated that total palm oil production in Kalimantan is predicted to remain relatively stable, with a slight increase by 2027. However, these predictions should be interpreted carefully because the historical data showed a sharp structural change after 2017, which may affect long-term forecasting stability.

The main contribution of this study is the development of a comparative panel time-series forecasting framework for regional palm oil production in Kalimantan. Unlike studies that rely on a single model or aggregate time-series data, this study compared multiple model families using the same experimental setting and evaluated their performance at both global and provincial levels. The findings provide methodological insight into the suitability of different forecasting models for heterogeneous regional production data.

From a practical perspective, the results can support regional agricultural planning, production monitoring, supply chain preparation, and policy-making in the palm oil sector. LSTM can be considered when the main objective is to reduce large forecasting errors, while Naive Forecast can be used as a strong and simple baseline for stable regional production patterns.

Future research should use longer historical data, include explanatory variables such as plantation area, rainfall, temperature, land suitability, fertilizer use, commodity prices, and policy factors, and apply rolling-origin validation for more robust evaluation. Further studies may also explore hybrid or ensemble models combining LSTM, XGBoost, and N-BEATS, as well as explainable artificial intelligence methods such as SHAP to identify the most influential factors affecting palm oil production forecasts.

References

[1] J. Supriatna, D. Djumarno, A. B. Saluy, and D. Kurniawan, "Sustainability Analysis of Smallholder Oil Palm Plantations in Several Provinces in Indonesia," *Sustainability*, vol. 16, no. 11, p. 4383, May 2024, doi: 10.3390/su16114383.

[2] L. Judijanto, "Beyond the plantation: Palm oil as a strategic lever for regional development and economic transformation growth in Indonesia," *Growth*, vol. 12, no. 1, pp. 25–32, Jun. 2025, doi: 10.20448/growth.v12i1.6842.

[3] C. L. Duta, J. Wuri, A. Budisusila, and L. B. Harnoto, "Analysis of the Role of Oil Palm Plantation Sub-Sector in Gross Regional Domestic Product of West Kalimantan Province," *Eduvest - Journal of Universal Studies*, vol. 5, no. 12, pp. 14737–14748, Dec. 2025, doi: 10.59188/eduvest.v5i12.52102.

[4] S. Rahmawati, "Provincial palm production analysis East Kalimantan 2016-2020," *ProBisnis : Jurnal Manajemen*, vol. 15, no. 1, pp. 71–76, Feb. 2024, doi: 10.62398/probis.v15i1.409.

[5] N. U. Khasanah, A. Suyatno, and M. Maswadi, "THE EFFECT OF OIL PALM PRODUCTIVITY ON THE GROSS REGIONAL DOMESTIC PRODUCT IN WEST KALIMANTAN PROVINCE," *SEPA: Jurnal Sosial Ekonomi Pertanian dan Agribisnis*, vol. 22, no. 2, p. 234, Sep. 2025, doi: 10.20961/sepa.v22i2.93044.

[6] L. N. Syahid, X. Luo, R. Zhao, and J. S. H. Lee, "Climate drives variation in remotely sensed palm oil yield in Indonesia and Malaysia," *Environmental Research Letters*, vol. 20, no. 4, p. 044016, Apr. 2025, doi: 10.1088/1748-9326/adbfac.

[7] J. Zhao, A. J. Elmore, J. S. H. Lee, I. Numata, X. Zhang, and M. A. Cochrane, "Replanting and yield increase strategies for alleviating the potential decline in palm oil production in Indonesia," *Agric. Syst.*, vol. 210, p. 103714, Aug. 2023, doi: 10.1016/j.agsy.2023.103714.

[8] M. Ben Ben Nasr and S. Ben Othman, "Time Series Forecasting of Suicide Mortality in Europe: A Comparative Performance Analysis of ARIMA, Holt-Winters Exponential Smoothing, and Naive Models (1950--2019)," *American Journal of Medical Science and Innovation*, vol. 5, no. 1, pp. 24–28, Jan. 2026, doi: 10.54536/ajmsi.v5i1.6363.

[9] N. Beck, J. Dovern, and S. Vogl, “Mind the naive forecast! a rigorous evaluation of forecasting models for time series with low predictability,” *Applied Intelligence*, vol. 55, no. 6, p. 395, Apr. 2025, doi: 10.1007/s10489-025-06268-w.

[10] L. W. Rizkallah, “Enhancing the performance of gradient boosting trees on regression problems,” *J. Big Data*, vol. 12, no. 1, p. 35, Feb. 2025, doi: 10.1186/s40537-025-01071-3.

[11] A. Karentia, F. Winaya, and D. Suhartono, “Hybrid approach sentiment analysis using Transformer-LSTM in the Indonesian language,” in *2024 7th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, Dec. 2024, pp. 754–758. doi: 10.1109/ISRITI64779.2024.10963365.

[12] M. Pirani, P. Thakkar, P. Jivrani, M. H. Bohara, and D. Garg, “A Comparative Analysis of ARIMA, GRU, LSTM and BiLSTM on Financial Time Series Forecasting,” in *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, IEEE, Apr. 2022, pp. 1–6. doi: 10.1109/ICDCECE53908.2022.9793213.

[13] T. Juhana, H. Yuliana, . Hendrawan, . Iskandar, and Y. Musashi, “Comparative Analysis of ARIMA, Prophet, and Glmnet for Long Term Evolution (LTE) Base Station Traffic Forecasting,” *Emerging Science Journal*, vol. 8, no. 6, pp. 2197–2217, Dec. 2024, doi: 10.28991/ESJ-2024-08-06-04.

[14] M. Usmani, Z. A. Memon, A. Zulfiqar, and R. Qureshi, “Preptimize: Automation of Time Series Data Preprocessing and Forecasting,” *Algorithms*, vol. 17, no. 8, p. 332, Aug. 2024, doi: 10.3390/a17080332.

[15] N. Beck, J. Dovern, and S. Vogl, “Mind the naive forecast! a rigorous evaluation of forecasting models for time series with low predictability,” *Applied Intelligence*, vol. 55, no. 6, p. 395, Apr. 2025, doi: 10.1007/s10489-025-06268-w.

[16] M. F. Rizvi, “ARIMA Model Time Series Forecasting,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 5, pp. 3782–3785, May 2024, doi: 10.22214/ijraset.2024.62416.

[17] N. Q. Mohamad Fozi and A. Abdul Aziz, “A NOVEL HYBRID HOLT INTEGRATED MOVING AVERAGE (HIMA) FOR CONSUMER PRICE INDEX PREDICTION,” *International Journal of Entrepreneurship and Management Practices*, vol. 7, no. 25, pp. 367–378, Jun. 2024, doi: 10.35631/IJEMP.725030.

[18] A. Choudhury, A. Mondal, and S. Sarkar, “Searches for the BSM scenarios at the LHC using decision tree-based machine learning algorithms: a comparative study and review of random forest, AdaBoost, XGBoost and LightGBM frameworks,” *Eur. Phys. J. Spec. Top.*, vol. 233, no. 15–16, pp. 2425–2463, Nov. 2024, doi: 10.1140/epjs/s11734-024-01308-x.

[19] K. Astudillo, M. Flores, M. Soliz, G. Ferreira, and J. Varela-Aldás, “A Hybrid GAS-ATT-LSTM Architecture for Predicting Non-Stationary Financial Time Series,” *Mathematics*, vol. 13, no. 14, p. 2300, Jul. 2025, doi: 10.3390/math13142300.

[20] M. Thor and Ł. Postek, “Gated recurrent unit network: A promising approach to corporate default prediction,” *J. Forecast.*, vol. 43, no. 5, pp. 1131–1152, Aug. 2024, doi: 10.1002/for.3057.

[21] L. F. R. Agottani, M. M. Girardi, W. R. de Souza Chaves, L. de Azevedo Takara, L. dos Santos Coelho, and V. C. Mariani, “Multi-step ahead wind power forecasting based on N-BEATS model,” in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, IEEE, Oct. 2024, pp. 1–5. doi: 10.1109/ISGTEUROPE62998.2024.10863789.

[22] A. Yunita *et al.*, “Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models,” *MethodsX*, vol. 15, p. 103462, Dec. 2025, doi: 10.1016/j.mex.2025.103462.

[23] S. Athghara and S. Arora, “A Comparative Analysis of Air Quality Index Forecasting Using LSTM, GRU and Multi-Variate NBEATS,” in *2025 10th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, Oct. 2025, pp. 2116–2122. doi: 10.1109/ICCES67310.2025.11337214.

[24] A. Yunita *et al.*, “Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models,” *MethodsX*, vol. 15, p. 103462, Dec. 2025, doi: 10.1016/j.mex.2025.103462.

[25] W. Li and K. L. E. Law, “Deep Learning Models for Time Series Forecasting: A Review,” *IEEE Access*, vol. 12, pp. 92306–92327, 2024, doi: 10.1109/ACCESS.2024.3422528.