Indonesian Journal of Electrical Engineering and Computer Science
Vol. 42, No. 2, May 2026, pp. 442~453
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v42.i2.pp442-453

Meta-stacking models for electricity load forecasting in West Java

Denanda Aufadlan Tsaqif, Bagus Sartono, Hari Wijayanto
School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

Article Info

Article history:

Received Dec 3, 2025
Revised Mar 7, 2026
Accepted May 26, 2026

Keywords:

CNN-BiGRU
CNN-BiLSTM
Electricity load forecasting
Meta-stacking
Windowed-XGBoost

ABSTRACT

Indonesia's electricity demand continues to increase due to population growth, urbanization, and industrial expansion, therefore making accurate load forecasting is essential to maintain supply-demand balance. However, electrical load demand in West Java has a complex pattern (seasonality, nonlinear behavior, weather variability, and holiday effects), which motivates the use of a meta-stacking approach to effectively capture such complexity. Previous research shows that meta-stacking outperforms individual models, but it fails to capture sudden changes and its performance consistency remains unclear. Therefore, this study proposes a meta-stacking framework for daily electricity load forecasting in West Java (2006-2023) that includes weather and holiday variables by combining CNN-BiLSTM, CNN-BiGRU, and Windowed-XGBoost forecasts through linear regression and evaluates its performance across five data-splitting scenarios and nine forecast horizons, which represents the main novelty in this research. Meta-stacking shows strong generalization across scenarios and strong long-term forecasting performance across horizons, while consistently providing a balanced trade-off between MAPE and trend accuracy, where the model trained on the longest historical dataset achieves the best performance with 1.89% MAPE and 86% trend accuracy. The proposed approach successfully captures seasonal and holiday-related load patterns, indicating its potential to support PLN in improving demand planning and operational decision-making.

Corresponding Author:

Denanda Aufadlan Tsaqif
School of Data Science, Mathematics, and Informatics, IPB University
16680 Bogor Regency, West Java, Indonesia
Email: denandaaufadlants@gmail.com

1. INTRODUCTION

Electric load represents the amount of electrical energy consumed by users over time and is categorized as time series data, which contain temporal dependencies and complex patterns such as trend, seasonality, and random fluctuations [1]. In Indonesia, electricity demand has been increasing rapidly due to population growth, urbanization, and industrial expansion, thus posing significant challenges for reliable supply planning and infrastructure development [2]. Forecasting plays an essential role in identifying these patterns to support decision-making in the energy sector [3]. Accurate electricity demand forecasting helps maintain the balance between supply and demand, prevent power disruptions, and improve distribution efficiency [4]. In regions like Bogor City, West Java that are characterized by high rainfall, dense population, and multiple external factors such as temperature, rainfall, and electricity tariffs, reliable forecasting is needed for effective energy planning [5], [6].

Traditional statistical models such as ARIMA offer interpretability but have limitations when handling multivariate and nonlinear relationships in forecasting [7]. Artificial intelligence models, including XGBoost, CNN, LSTM, and GRU, overcome these challenges by capturing complex and nonlinear temporal patterns [8], [9]. LSTM effectively learns long-term dependencies, while GRU achieves higher computational efficiency for short-term dependencies [10]. To further improve forecasting performance, hybrid approaches have been proposed. One of the most promising approaches is meta-stacking, which integrates multiple base-models and uses a meta-learner to combine their outputs, often result in superior forecast accuracy [11], [12].

Previous study [10] used meta-stacking to forecast electricity load in Australia and Spain. The model consists of LSTM, Bi-LSTM, XGBoost, and stacked LSTM as base-learners and lasso regression as meta-learner. The proposed model was shown to outperform individual models in terms of MAPE, but the proposed models fail to forecast the minimum load values and fail to anticipate sudden changes which influenced by calendar factor. Multiple LSTM variants in the base-models result in similar forecasting results between each variant, thus advantages captured by the meta-learner is suboptimal. Therefore, using more varied base-models can improve the meta-learner performance in exploiting the advantages of complementary models. Linear regression is the most effective approach to exploit the advantages of complementary information [13]. Also, the trend accuracy is not considered in the research, so it cannot be evaluated how the proposed model accurately forecast the trend, especially in short-terms. The proposed model only tested using two different location each with only one scenario of training and testing data, so its evaluation is limited in terms of model consistency and does not consider the effect of very long-time information.

This study proposes a meta-stacking model that integrates Windowed-XGBoost, CNN-BiGRU, and CNN-BiLSTM as base-models, with linear regression as the meta-learner. Windowed-XGBoost is selected for its simplicity and efficiency [14], while CNN-BiGRU and CNN-BiLSTM effectively capture spatial and temporal dependencies bidirectionally [15], [16]. This study also explicitly considered calendar factors. The proposed model aims to achieve accurate and consistent forecasts across different data-splitting scenarios in terms of MAPE and trend accuracy. The results are expected to provide insights for PLN in optimizing electricity supply planning in West Java and contribute to the application of hybrid methods in energy demand forecasting.

The objectives of this research are to model and evaluate the proposed model for forecasting electricity load in West Java, to test its performance and compare it with individual models across different scenarios and forecast horizons, and to identify the best-performing model while evaluating its ability to accurately forecast the electricity load data in West Java. This study seeks to address the following research questions: i) Can the meta-stacking model outperform individual models across different data-splitting scenarios and forecast horizons? (ii) To what extent can the proposed model accurately forecast electricity load in West Java, particularly those driven by calendar effects and sudden demand changes?

2. METHOD

2.1. Data

This study used electricity, weather, and calendar data in West Java, consisting of daily observations from January 1, 2006, to December 31, 2023, with a total of 6541 data points. The electricity data consist of 30-minute load measurements in megawatts (MW), which are later aggregated into daily values through summation and used as the forecasting target. The dataset was obtained from PLN UIP2B JAMALI. In addition to the electricity data, weather data were used as external features to support the forecasting analysis. The weather data were obtained from the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG) website from stations: Bandung Geophysics Station, West Java Climatology Station (Bogor City), Citeko Meteorology Station (Bogor Regency), Kertajati Meteorology Station, and Penggung Meteorology Post (Majalengka), and the measurements from these stations were averaged for each variable to represent the overall climatic conditions of West Java. The weather variables include average temperature, rainfall intensity, duration of sunshine, humidity, and wind speed. Another external feature used in this research is calendar data, which identifies national and religious holidays such as Independence Day, Eid al-Adha, Eid al-Fitr, and New Year, since electricity load is strongly affected by holidays [17], [18]. For example, in the days leading up to Eid al-Fitr, electricity usage consistently drops because many factories are shut down [19].

2.2. Framework of the proposed model

The research workflow illustrated in Figure 1 starts with data exploration, followed by scenario-based data splitting into training and testing sets. The data is then split into five scenarios based on Table 1. For each scenario, the training data are divided into two phases, where the first training dataset is pre-processed and used to build individual base-models, while the second training dataset, which is obtained by splitting the initial training data before pre-processing, is used for meta-stacking modelling. The data validation is used to generate out-of-fold (OOF) predictions as part of the meta-stacking mechanism. This whole process is repeated until all five scenarios have been completed. Finally, the overall model performance is evaluated to obtain the final results of the forecasting framework, as shown in Figure 1. This entire modelling flow is repeated for all five scenarios. Individual models and meta-stacking model are then evaluated using MAPE and trend accuracy using test data and are evaluated on different horizon which can be seen on

Table *2*.

The use of multiple data splitting scenarios aims to examine the consistency and robustness of the proposed models under different training lengths based on Table 1. By extending the starting year of the training data from 2018 back to 2006, this scheme allows the analysis of whether a longer historical dataset leads to better forecasting performance. Scenario 5 is specifically designed to evaluate how the models perform when trained on a shorter and more limited dataset.

The testing horizon scenarios in

Table *2* are designed to evaluate model performance across different forecasting lengths, ranging from short-term to long-term predictions. This scheme allows the analysis of how model performed when forecasting for 1 week, several weeks, monthly, and quarterly horizons.

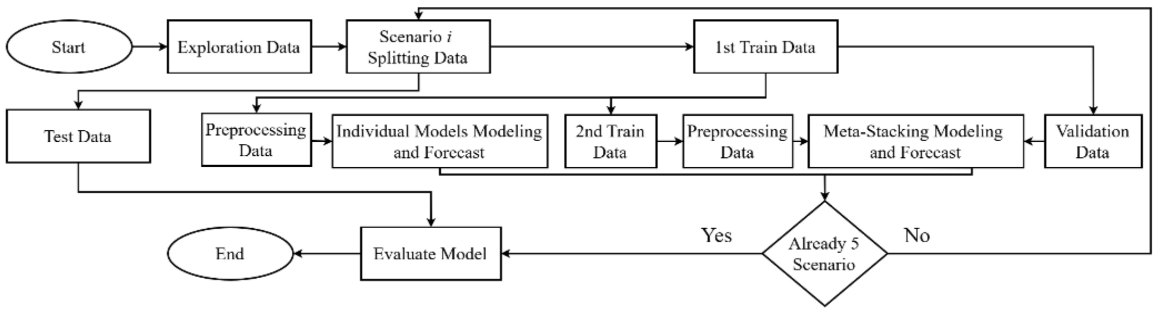


Figure 1. Framework of the proposed model

Table 1. Splitting data scenarios on West Java electricity load

Scenario	Training data	Testing data
1	01/01/2006 – 31/12/2022	01/01/2023 – 31/12/2023
2	01/01/2009 – 31/12/2022	01/01/2023 – 31/12/2023
3	01/01/2012 – 31/12/2022	01/01/2023 – 31/12/2023
4	01/01/2015 – 31/12/2022	01/01/2023 – 31/12/2023
5	01/01/2018 – 31/12/2022	01/01/2023 – 31/12/2023

Table 2. Testing data horizon scenarios

Scenario	Horizon	Testing data
1	1 Week	01/01/2023 – 08/01/2023
2	2 Week	01/01/2023 – 15/01/2023
3	3 Week	01/01/2023 – 22/01/2023
4	4 Week	01/01/2023 – 29/01/2023
5	1 Month	01/01/2023 – 31/01/2023
6	2 Month	01/01/2023 – 28/02/2023
7	1 Quarter	01/01/2023 – 31/03/2023
8	2 Quarter	01/01/2023 – 30/06/2023
9	3 Quarter	01/01/2023 – 30/09/2023

2.3. Data preprocessing

Both training datasets undergo a preprocessing process, as illustrated in Figure 2, which is differentiated based on the model type: deep learning (CNN-BiLSTM and CNN-BiGRU) and Windowed-XGBoost. For the deep learning models, one-hot encoding is applied before data splitting, followed by Min-

Max normalization and window building after splitting. Window building transforms the time series into sequential input-output pairs. This research using a window size of 31 days to capture full monthly seasonal patterns. For Windowed-XGBoost, preprocessing consists of data splitting and significant lag extraction for the load variable using the partial autocorrelation function (PACF) plot. The use of the PACF plot provides statistical justification for selecting relevant lag terms by identifying significant partial autocorrelations, ensuring that only informative temporal dependencies are included in the model [20].

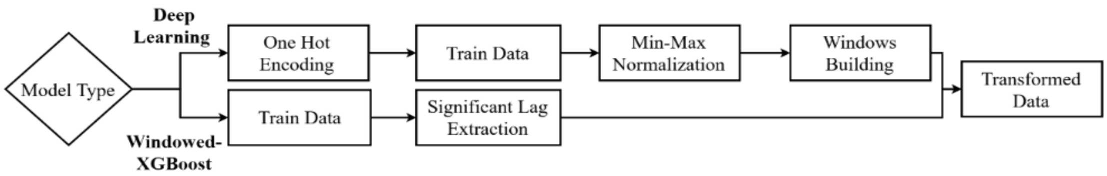


Figure 2. Preprocessing scheme for deep learning and Windowed-XGBoost

2.4. Modeling scheme

The whole modeling mechanism is illustrated in Figure 3. Each scenario has a different starting point for the training data, where the dataset is divided into two different training sets. The first training set is used for individual models modeling, including preprocessing, hyperparameter tuning, and generating forecasts. In this study, the individual models, which later act as base-models in the meta-stacking framework, consist of CNN-BiLSTM, CNN-BiGRU, and Windowed-XGBoost. The optimal hyperparameter configuration obtained from this stage is then reused in the second stage, where the second training set is used for meta-stacking modeling by re-training these base-models without further tuning, and their predictions are combined using a linear regression model as the meta-learner.

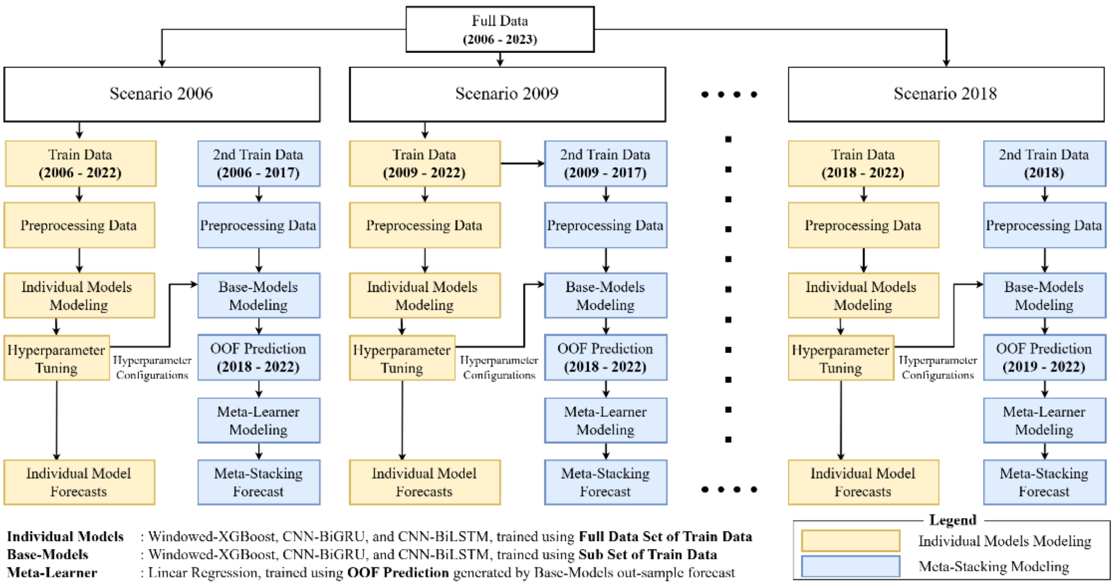


Figure 3. Modeling scheme

Hyperparameter optimization is conducted using Bayesian optimization to efficiently explore the defined search space. The CNN component applies one or two convolutional layers with filters ranging from 16 to 128 and kernel sizes of 2, 3, or 5, and activation functions including ReLU, tanh, or linear. The recurrent component consists of one or two bidirectional layers (BiGRU or BiLSTM) with 16 to 128 neurons per layer and activation functions of tanh, ReLU, or linear. Model optimization is performed using Adam or RMSprop with a learning rate sampled within 10^{-4} to 10^{-2}, and training is conducted using batch sizes of 16, 32, or 64. For the Windowed-XGBoost model, the architecture is based on gradient boosting decision trees with lag-based input features. The number of trees (n_estimators) ranges from 50 to 1000, with maximum tree depth between 3 and 15. The boosting learning rate is searched within 0.001 to 0.3, while sampling parameters include subsample within 0.6 to 1.0 and colsample_bytree within 0.5 to 1.0. Regularization parameters include gamma within 0 to 5, min_child_weight within 1 to 15, L2 regularization (λ) within 1e-3 to 100, and L1 regularization (α) within 1e-3 to 100 to control model complexity and prevent overfitting.

Wolpert [21] introduced meta-stacking, a hybrid model approach that combines multiple models, known as base-learners, through another model called as meta-learner. Meta-stacking consists of two levels level 0, where base-models are trained in parallel using the second training data, and level 1, where level 0 forecasts are used as input for the meta-learner. To ensure unbiased meta-learner input, OOF predictions are generated through an iterative mechanism, as illustrated in Figure 4. The second training data are divided into several iterations, where each iteration uses one subset as training data while the remaining data are used for validation data. The validation results from Iteration 1 to Iteration 5 are combined to form OOF predictions across the entire second training data. These OOF predictions are then used as inputs to train the meta-learner, ensuring that each prediction is generated from a model that does not directly train on the same validation observations [20]. Meta-stacking approach allows the base-models to compensate for each other's limitations while leveraging their individual strengths [22], [23].

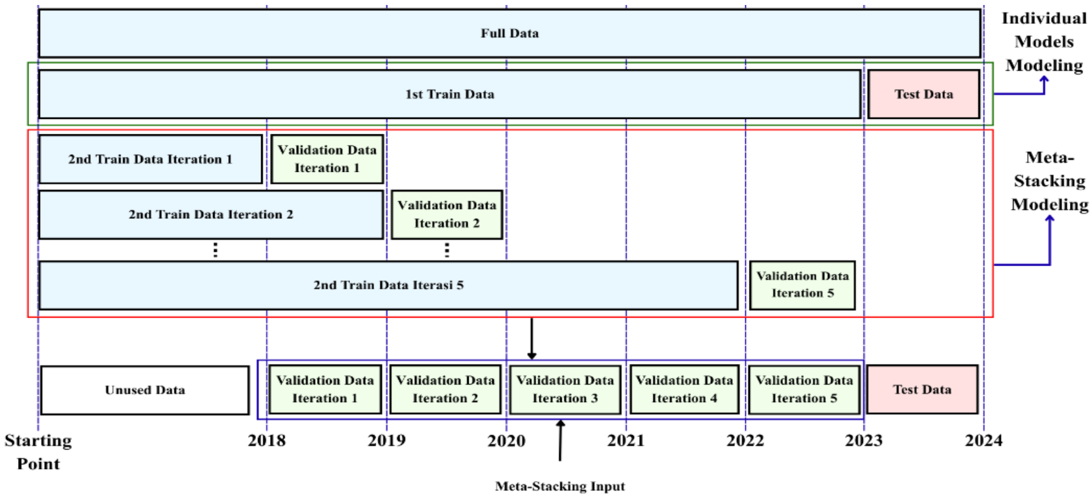


Figure 4. Out-of-fold predictions scheme

2.5. Model evaluation scheme

Model evaluation is conducted to assess the performance of meta-stacking compared to individual models, as illustrated in Figure 5. Forecasts are generated for all scenarios using data up to 2022 for modeling, while the year 2023 is used as the test dataset to evaluate performance on unseen data. The test data are further divided into multiple forecast horizons based on

Table *2*. The evaluation includes generalization analysis by comparing train and test accuracy, as well as performance assessment on the test data using MAPE and trend accuracy across different horizons. Finally, the best model is determined based on its overall forecasting performance under these evaluation criteria.

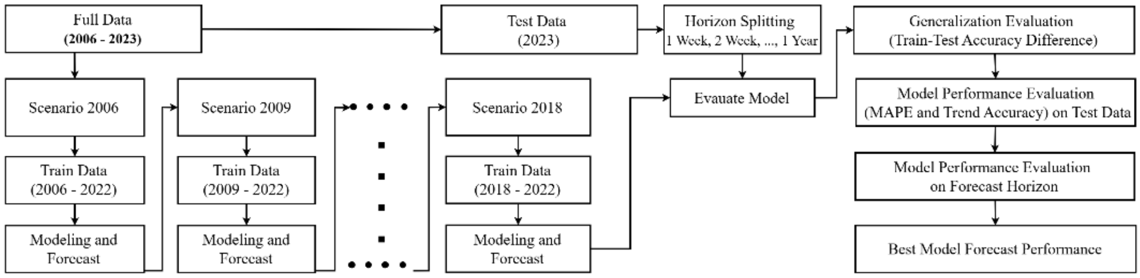


Figure 5. Model evaluation scheme

3. RESULTS AND DISCUSSION

In this section, we present an evaluation of the proposed meta-stacking model for electricity load forecasting in West Java Province. The model's performance is compared against three individual base-learners, which are Windowed-XGBoost, CNN-BiGRU, and CNN-BiLSTM. The model's also evaluated on five different scenarios.

3.1. Performance evaluation metrics

The metrics used in this study are Mean Absolute Percentage Error (MAPE) and trend accuracy. These metrics clearly show the differences between the predicted values and the actual values. MAPE is formulated as follows [24]:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\left(X_{act}(i)-X_{pred}(i)\right)}{X_{act}(i)}\right| \quad (1)$$

where $X_{act}(i)$ is the electricity load in period i, $X_{pred}(i)$ is the forecasted electricity in period i. The confusion matrix is used to measure classification accuracy, calculated as [25]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (2)$$

where trends are considered upward (Positive) if $y_{t+1} > y_t$ and downward (Negative) if $y_{t+1} < y_t$. The counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are obtained by comparing the direction of the actual and predicted trends.

3.2. Result

Before modelling the West Java electrical load data, an exploratory analysis is conducted to get a clear understanding of the characteristics and patterns within the data. Based on

Figure *6*, in the longer-term, the west java electrical load data from 2006 until 2023 has positive trends, which is caused by the more advanced technology requires more electric and human population increase [26]. There's also a repeated significant drop in the data, where the most significant drop is caused by Eid al-Fitr and Eid al-Adha. Also, others holiday such as New Year led to a significant drop. This is caused by many factories closed due to major holidays [19]. Covid-19 incident also led to a significant drop in the data which can be seen on the period of 2021. Based on these findings indicate that the electrical load data is driven by global and local event.

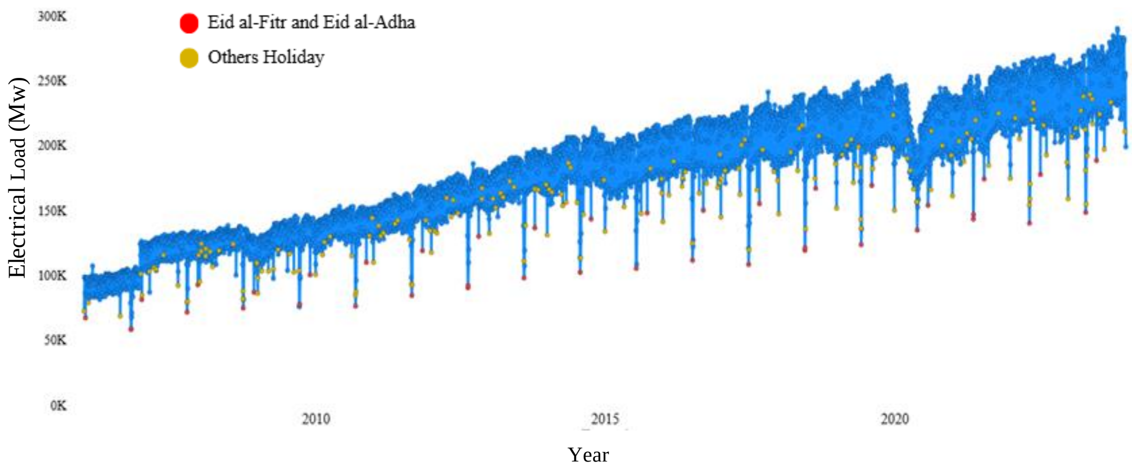


Figure 6. West Java electrical load data pattern from January 1, 2006 until December 31, 2023

Prior to analysis of the forecasting models, an initial analysis was performed to identify seasonal pattern in the electrical data load. The results indicate the data has 7-days seasonality autocorrelation pattern, as shown in

Figure *7*. We analyzed the forecasted electricity load in West Java Province across different scenarios to evaluate whether the proposed meta-stacking model outperforms the individual models.

Figure 8 presents the distribution of model performance across five scenarios using training and testing data, allowing evaluation on generalization capability.

Figure 8(a) illustrates the distribution of MAPE for each model. The meta-stacking model shows a more stable distribution and a smaller gap between training and testing errors, indicating better generalization.

Figure 8(b) presents the distribution of trend accuracy. The meta-stacking model maintains consistent trend performance across scenarios, demonstrating its robustness compared to the individual models. The gap between training and testing performance is relatively small, indicating good generalization ability compared to individual models, which tend to exhibit larger discrepancies between train and test results.

Error! Reference source not found. presents the comparison of testing performance across scenarios, highlighting both error magnitude and trend accuracy for each model. This figure aims to show how the models perform and their consistency across scenarios. **Error! Reference source not found.**(a) displays the testing MAPE, while **Error! Reference source not found.**(b) displays the testing trend accuracy. Based on **Error! Reference source not found.**(a), meta-stacking model achieves the lowest MAPE in scenarios 2006, 2015, and 2018. CNN-BiLSTM performs slightly better in 2009 and 2012, but shows relatively low trend accuracy. Based on **Error! Reference source not found.**(b) meta-stacking has the highest trend accuracy in most scenarios (2006, 2009, 2012, and 2015), whereas Windowed-XGBoost shows the highest trend accuracy in 2018 but with relatively high MAPE. Overall, meta-stacking provides a better balance between low forecasting error and consistent trend accuracy across scenarios.

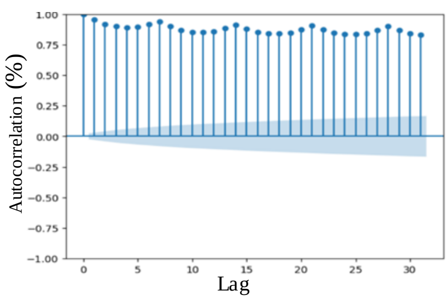
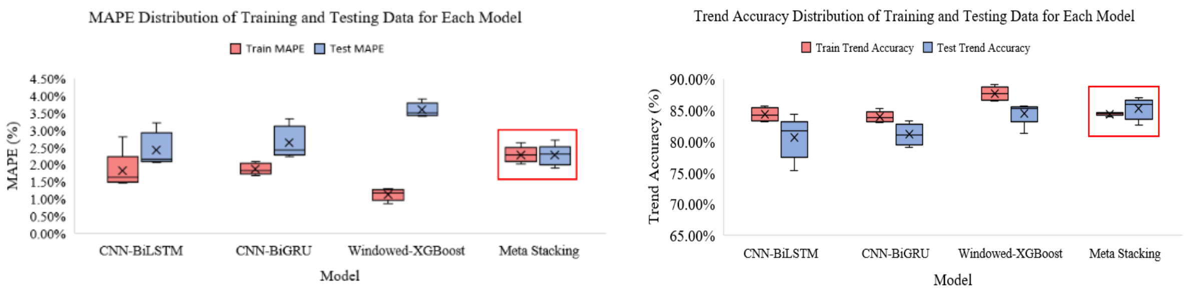
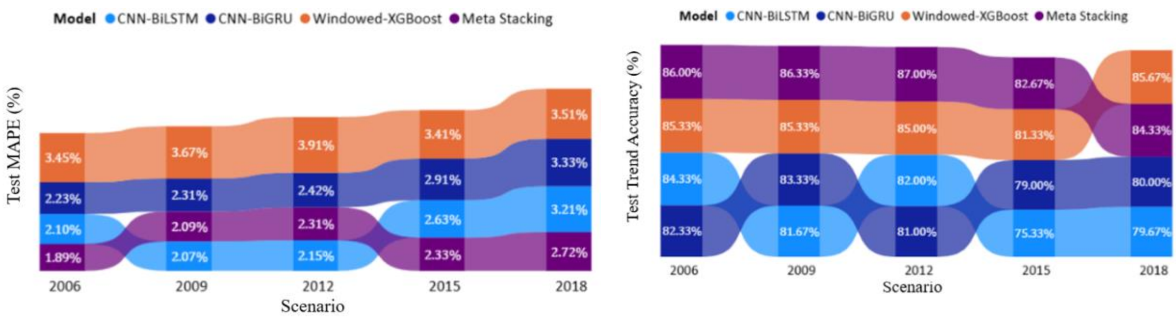


Figure 7. West Java electricity load autocorrelation pattern



(a) (b)

Figure 8. Distribution of (a) MAPE and (b) trend accuracy of training and testing data for each model



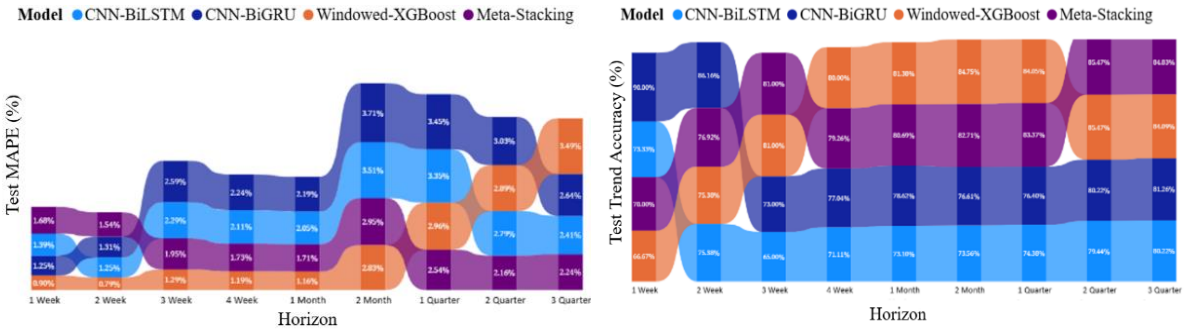
(a) (b)

Figure 9. Comparison of (a) MAPE and (b) trend accuracy of testing data for each scenario and model

Figure 10 presents the comparison of average testing performance across different forecast horizons, ranging from short-term (1-4 weeks) to longer-term horizons (up to 3 quarters). This figure aims to evaluate how each model performed as the forecasting length increases and to assess performance stability. Figure 10(a) shows the average MAPE across horizons, indicating that meta-stacking consistently achieves the lowest error for longer horizons (1 to 3 quarters), although it does not always dominate at very short horizons (1-2 weeks). Figure 10(b) illustrates the average trend accuracy, where meta-stacking generally performs better at longer horizons, while Windowed-XGBoost and CNN-BiGRU occasionally obtain the highest scores at certain shorter horizons. Overall, these results suggest that meta-stacking is more reliable for long-term forecasting, while Windowed-XGBoost may be a better choice for short-term prediction, further analysis is needed to check whether the model generalize well on the new data or not.

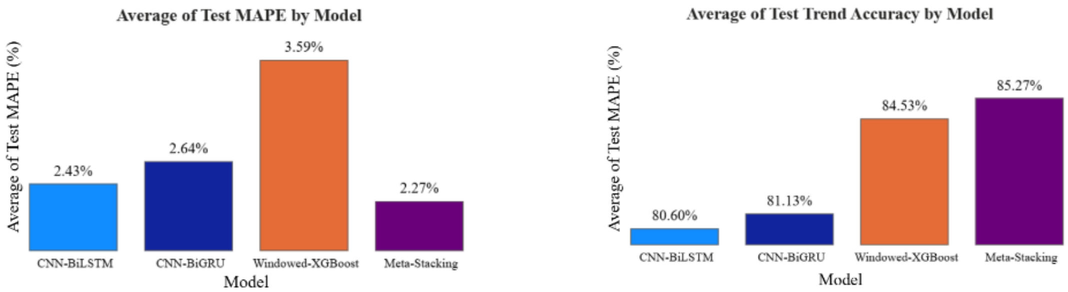
Figure 11 provides baseline comparison by averaging testing performance of all models across the five scenarios. Based on Figure 11(a), meta-stacking achieves the lowest average test MAPE (2.27%), while Figure 11(b) indicates that meta-stacking achieves the highest average test trend accuracy (85.27%), outperforming the individual models in both metrics. While Windowed-XGBoost shows relatively high average test trend accuracy, it also achieves higher average test MAPE. On the contrary, CNN-BiGRU and CNN-BiLSTM achieves low average test MAPE, but achieves low average test trend accuracy. Overall, this result confirms that meta-stacking provides the balanced tradeoff between MAPE and trend accuracy, and superior performance among the individual models. Summary of the research results is presented on the

Table 3.



(a) (b)

Figure 10. Average (a) MAPE and (b) trend accuracy of 5 scenarios for each forecast horizon and model



(a) (b)

Figure 11. Average of (a) test MAPE and (b) test trend accuracy by model

Table 3. Summary of the research results

Model	Average test MAPE	Average test trend accuracy	Strengths	Weaknesses
CNN-BiLSTM	2.43%	80.60%	Low MAPE, slightly competitive with meta-stacking in MAPE	Lowest trend accuracy
CNN-BiGRU	2.64%	81.13%	Low MAPE	Less competitive in MAPE
Windowed-XGBoost	3.59%	84.53%	High trend accuracy, slightly competitive with meta-stacking in trend accuracy	High MAPE, weaker generalization
Meta-stacking	2.27%	85.27%	Lowest MAPE and Trend Accuracy, robust generalization	High computational

Based on the research results, meta-stacking can leverage Windowed-XGBoost strength on high trend accuracy and deep learning models strength on low MAPE, thus indicating meta-stacking successfully leverage base-models strength, while overcoming both weaknesses. Meta-stacking is a better model because of its ability in effectively generalize on the new data, its superiority against individual models in most scenarios while maintaining balance tradeoff between MAPE and trend accuracy, and its superiority in long-term forecasting. Based on Table 4, meta-stacking in scenario of 2006 has the lowest MAPE and the 3rd highest trend accuracy, which is the best model in this study. Figure 12 illustrates the train-test performance of the meta-stacking model across scenarios, while Figure 12(a) based on MAPE and Figure 12(b) based on trend accuracy). Figure 12(a) shows that the longer the dataset, the better meta-stacking performs in terms of MAPE. On the contrary, Figure 12(b) shows a different behavior for trend accuracy, which fluctuates across scenarios and does not consistently increase with more data. The consistent improvement of MAPE becomes the basis for determining the best-performing model in this study. The accuracy of all models across 5 scenarios is explained in Table 4.

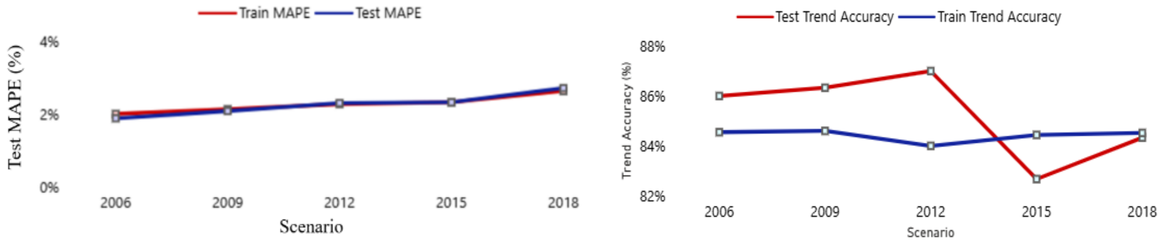
Figure 133 presents the forecasting performance of the meta-stacking model on the 2023 testing data, including the actual versus predicted values in Figure 133(a) and the residual analysis in Figure 133(b). As shown in Figure 133(a), the predicted series closely follows the actual load across most periods, capturing both upward and downward movements with good consistency. The model remains stable during periods of sudden spikes or drops, indicating strong generalization on unseen data and reliable performance for operational forecasting.

Despite the strong overall predictive performance observed in Figure 133(a), residual gaps are still observed, as illustrated in Figure 133(b). The largest errors tend to occur during national holidays and joint leave periods. Although holiday indicators were included as model features, uncommon holidays or major holidays still generate higher residuals due to sudden shifts in electricity demand. Overall, the results indicate that the proposed meta-stacking model achieves a more balanced performance between MAPE and trend accuracy across multiple data-splitting scenarios compared to the individual models. The model trained under the 2006 scenario demonstrates the strongest overall performance in terms of forecasting accuracy and generalization on unseen data. Making it reliable to support PLN operational needs.

The proposed framework presents potential scalability for real-time forecasting in PLN operational. The base-models are trained using historical data, and the meta-learner combines their forecasts through a simple regression approach. As a result, the prediction process remains computationally efficient and can generate daily load forecasts with minimal processing time. This makes the proposed framework suitable for integration into PLN's operational planning.

Table 4. MAPE and trend accuracy of all models across five scenarios (train and test performance)

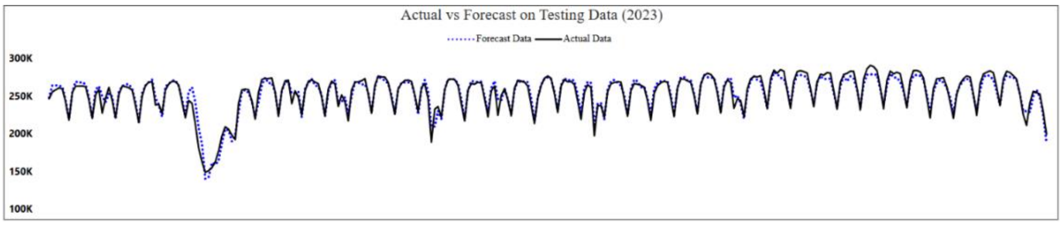
Scenario	Type	CNN-BiLSTM		CNN-BiGRU		Windowed-XGBoost		Meta-stacking	
		MAPE	Trend accuracy	MAPE	Trend accuracy	MAPE	Trend accuracy	MAPE	Trend accuracy
2006	Train	1.63%	84.16%	1.81%	83.18%	1.26%	86.74%	2.01%	84.55%
	Test	2.10%	84.33%	2.23%	82.33%	3.45%	85.33%	1.89%	86.00%
2009	Train	1.66%	83.18%	1.77%	85.32%	1.28%	86.41%	2.14%	84.60%
	Test	2.07%	81.67%	2.31%	83.33%	3.67%	85.33%	2.09%	86.33%
2012	Train	1.52%	85.10%	1.67%	83.89%	1.17%	87.73%	2.28%	84.00%
	Test	2.15%	82.00%	2.42%	81.00%	3.91%	85.00%	2.31%	87.00%
2015	Train	1.46%	85.70%	2.00%	83.07%	1.07%	88.38%	2.33%	84.44%
	Test	2.63%	75.33%	2.91%	79.00%	3.41%	81.33%	2.33%	82.67%
2018	Train	2.80%	83.36%	2.09%	84.11%	0.86%	89.18%	2.64%	84.52%
	Test	3.21%	79.67%	3.33%	80.00%	3.51%	85.67%	2.72%	84.33%



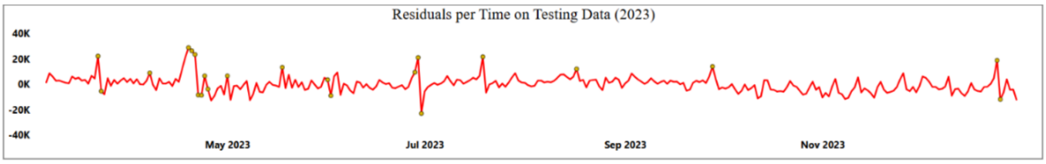
(a) (b)

Figure 12. Train-test (a) MAPE and (b) trend accuracy based on train-test dataset of the meta-stacking model across all scenarios

However residual errors still occur during uncommon and major holiday periods, such as Chinese New Year and Eid al-Fitr. Nevertheless, the model effectively captures both short-term and long-term variations in electricity load. It is important to note that the current meta-learner is trained only on the forecasts generated by the base-models, without explicitly incorporating calendar-related features. Future improvements may consider integrating calendar factors into the meta-learner to better capture demand shifts driven by holiday effects.



(a)



(b)

Figure 13. Comparison of (a) actual vs. forecasted load and (b) residuals with holiday markers on the 2023 testing data

4. CONCLUSION

The meta-stacking model outperforms the individual models in most scenarios based on both MAPE and trend accuracy and shows superior performance for long-term forecasts (1 quarter and beyond). The best-performing model, trained on the longest dataset (starting from 2006), achieves a MAPE of 1.89% and a trend accuracy of 86%, successfully forecasting the following year and capturing seasonal electricity load patterns. These results highlight the scientific contribution of this study in establishing a multi-scenario data-splitting evaluation that validates the stability of meta-stacking across different data conditions and forecast horizons.

However, limitations still occur during major holidays such as Chinese New Year and Eid al-Fitr, suggesting that explicitly including calendar features into the meta-learner could improve its ability to capture holiday-related demand shifts. Future work may consider validating the model on larger datasets across multiple provinces to assess meta-stacking scalability, as well as benchmarking the proposed approach against similar models in other countries. Including additional exogenous variables such as electricity tariffs, the number of registered PLN customers, industrial activity indicators, or economic growth may further enhance the model's ability to capture structural demand changes. Model training can also be refined by integrating trend accuracy into the optimization process, allowing the model to simultaneously optimize forecasting error and trend accuracy. Additionally, improving model interpretability through SHAP analysis would provide deeper insights into feature contributions at both base-model and meta-learner levels, while careful consideration of the trade-off between forecasting accuracy and computational resources remains important for practical deployment.

FUNDING INFORMATION

This research was supported by the Ministry of Education, Culture, Research, and Technology (Kemendiktisaintek) of Indonesia through research assistance. No additional external funding was received for this study.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Denanda Aufadlan Tsaqif	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Bagus Sartono	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
Hari Wijayanto		✓		✓	✓	✓		✓		✓		✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known financial, personal, academic, or professional conflicts of interest that could have influenced the work reported in this manuscript. There are no political, religious, or ideological competing interests associated with the preparation, analysis, or publication of this study. Authors state no conflict of interest.

DATA AVAILABILITY

The electricity load data used in this study were obtained from PLN UIP2B JAMALI and are subject to confidentiality restrictions. These data are not publicly available and may be accessed only with permission from the respective data provider. The weather data supporting this research were obtained from the BMKG (Indonesian Meteorology, Climatology, and Geophysics Agency) through publicly accessible sources on its official website (https://dataonline.bmkg.go.id/). Calendar data used in this study are publicly available from national and religious holiday information published by the Government of Indonesia. Additional derived data or processed datasets are available from the corresponding author upon reasonable request.

REFERENCES

[1] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*, 3rd ed. in Springer Texts in Statistics. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-29854-2.

[2] M. A. McNeil, N. Karali, and V. Letschert, "Forecasting Indonesia's electricity load through 2030 and peak demand reductions from appliance and lighting efficiency," *Energy for Sustainable Development*, vol. 49, pp. 65–77, Apr. 2019, doi: 10.1016/j.esd.2019.01.001.

[3] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 2nd ed. Melbourne: OTexts, 2018.

[4] H. Habbak, M. Mahmoud, K. Metwally, M. M. Fouda, and M. I. Ibrahem, "Load forecasting techniques and their applications in smart grids," *Energies*, vol. 16, no. 3, p. 1480, Feb. 2023, doi: 10.3390/en16031480.

[5] J. Son, J. Cha, H. Kim, and Y. M. Wi, "Day-ahead short-term load forecasting for holidays based on modification of similar days' load profiles," *IEEE Access*, vol. 10, pp. 17864–17880, 2022, doi: 10.1109/ACCESS.2022.3150344.

[6] C. Yin and S. Mao, "Fractional multivariate grey Bernoulli model combined with improved grey wolf algorithm: application in short-term power load forecasting," *Energy*, vol. 269, p. 126844, Apr. 2023, doi: 10.1016/j.energy.2023.126844.

[7] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting," *Energy*, vol. 214, p. 118874, Jan. 2021, doi: 10.1016/j.energy.2020.118874.

[8] D. Solyali, "A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus," *Sustainability*, vol. 12, no. 9, p. 3612, Apr. 2020, doi: 10.3390/su12093612.

[9] H. Rahman, M. U. Ahmed, S. Barua, P. Funk, and S. Begum, "Vision-based driver's cognitive load classification considering eye movement using machine learning and deep learning," *Sensors*, vol. 21, no. 23, p. 8019, Nov. 2021, doi: 10.3390/s21238019.

[10] F. Guo *et al.*, "A hybrid stacking model for enhanced short-term load forecasting," *Electronics*, vol. 13, no. 14, p. 2719, Jul. 2024, doi: 10.3390/electronics13142719.

[11] G. Dudek, "Combining forecasts using meta-learning: a comparative study for complex seasonality," in *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, Oct. 2023, pp. 1–10. doi: 10.1109/DSAA60987.2023.10302585.

[12] Y. Zhang, J. Ma, S. Liang, X. Li, and J. Liu, "A stacking ensemble algorithm for improving the biases of forest aboveground biomass estimations from multiple remotely sensed datasets," *GIScience & Remote Sensing*, vol. 59, no. 1, pp. 234–249, Dec. 2022, doi: 10.1080/15481603.2021.2023842.

[13] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2012.

[14] X. Zhao, Q. Li, W. Xue, Y. Zhao, H. Zhao, and S. Guo, "Research on ultra-short-term load forecasting based on real-time electricity price and window-based XGBoost model," *Energies*, vol. 15, no. 19, p. 7367, Oct. 2022, doi: 10.3390/en15197367.

[15] N. Mughees, S. A. Mohsin, A. Mughees, and A. Mughees, "Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting," *Expert Systems with Applications*, vol. 175, p. 114844, Aug. 2021, doi: 10.1016/j.eswa.2021.114844.

[16] B. Qin *et al.*, "A hybrid deep learning model for short-term load forecasting of distribution networks integrating the channel attention mechanism," *IET Generation, Transmission & Distribution*, vol. 18, no. 9, pp. 1770–1784, May 2024, doi: 10.1049/gtd2.13142.

[17] F. Rodrigues, C. Cardeira, J. M. F. Calado, and R. Melicio, "Short-term load forecasting of electricity demand for the residential sector based on modelling techniques: A systematic review," *Energies*, vol. 16, no. 10, p. 4098, May 2023, doi: 10.3390/en16104098.

[18] F. Ziel, "Modeling public holidays in load forecasting: a German case study," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 2, pp. 191–207, Mar. 2018, doi: 10.1007/s40565-018-0385-5.

[19] S. H. Ramalan, "Factories closed, PLN predicts electricity demand will drop by 30% during the Eid holiday," SindoNews. Accessed: Nov. 25, 2025. [Online]. Available: https://ekbis.sindonews.com/read/1545417/34/pabrik-pabrik-tutup-pln-prediksi-beban-listrik-turun-30-saat-libur-lebaran-1742461380

[20] S. M. Dadan Kusnandar, Suaibah, "Double seasonal autoregressive integrated moving average modeling in electricity load usage forecasting," *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 9, no. 1, pp. 21–30, Jan. 2020, doi: 10.26418/bbimst.v9i1.38028.

[21] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, Jan. 1992, doi: 10.1016/S0893-6080(05)80023-1.

[22] L. Nguyen Van and G. Lee, "Optimizing stacked ensemble machine learning models for accurate wildfire severity mapping," *Remote Sensing*, vol. 17, no. 5, p. 854, Feb. 2025, doi: 10.3390/rs17050854.

[23] M. H. Ganjgah, P. Azhdari, and K. F. Vajargah, "A hybrid stacking ensemble approach combining fuzzy logistic regression and fuzzy neural networks for enhanced COVID-19 diagnostic accuracy," *Discover Applied Sciences*, vol. 7, no. 10, p. 1106, Sep. 2025, doi: 10.1007/s42452-025-07456-6.

[24] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *International Journal of Forecasting*, vol. 32, no. 3, pp. 669–679, Jul. 2016, doi: 10.1016/j.ijforecast.2015.12.003.

[25] S. Sathyanarayanan, "Confusion matrix-based performance evaluation metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.

[26] I. F. U. Muzayanah, H. H. Lean, D. Hartono, K. D. Indraswari, and R. Partama, "Population density and energy consumption: a study in Indonesian provinces," *Heliyon*, vol. 8, no. 9, p. e10634, Sep. 2022, doi: 10.1016/j.heliyon.2022.e10634.

BIOGRAPHIES OF AUTHORS

Denanda Aufadlan Tsaqif, S.Mat. graduated with a bachelor's degree in Mathematics at IPB University and continued his master's studies in Statistics and Data Science at IPB University. His academic interests include statistical modeling, time series forecasting, machine learning, and data-driven decision-making. Throughout his graduate study, he has applied quantitative methods to real problem, which is electricity load forecasting, hybrid meta-stacking approach, and with exogenous feature for predictive modeling. He has also gained hands-on experience in data engineering and analytics through industry-based projects. His research contributes to developing practical and reliable forecasting methods by combining deep learning models, ensemble-based approaches, and relevant external features to better capture real-world patterns in electricity load data. He can be contacted at email: denzzdenanda@apps.ipb.ac.id.

Dr. Bagus Sartono, M.Si. is a lecturer in the School of Data Science, Mathematics, and Informatics at IPB University. He earned his bachelor's and master's degrees in Statistics at IPB University, and followed by doctoral studies in Statistics with specialization in Applied Economics at Universiteit Antwerpen, Belgium. His research interests include applied statistics, statistical modelling, multivariate data analysis, data mining, and predictive analytics as evidenced by his extensive publication record spanning regression modeling, mixed models, classification, machine learning integration, and small-area estimation techniques. He has contributed to various empirical studies from public health to environmental data, poverty estimation, climate downscaling, and social-economics demonstrating versatility in applying statistical methodologies to interdisciplinary problems. Besides his academic duties, he is also active as a consultant and instructor in statistical methods, research design, and data analysis for various institutions and governmental bodies in Indonesia. He can be contacted at email: bagusco@apps.ipb.ac.id.

Prof. Dr. Ir. Hari Wijayanto, M. Si. is a full professor in the School of Data Science, Mathematics, and Informatics at IPB University, holding the academic rank of "Professor" in the field of Survey Analysis and Design. He earned his bachelor's and master's degrees in Statistics at IPB University, and followed by doctoral studies in Statistics at IPB University. Over his academic career, he has developed expertise in statistical modeling, mixed models, survey methodology, multivariate statistics, spatial statistics, and applied data analysis across sectors such as agriculture, economics, public health, and social sciences. He has authored/co-authored more than 80 peer-reviewed works, indicating a broad and sustained contribution to statistical research and methodology. In addition to academic duties, he is actively involved in curriculum development, teaching, and mentoring at IPB, contributing both to academic growth and practical applications of statistics in Indonesia. He can be contacted at email: hari@apps.ipb.ac.id.