JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9980
Hal 1412-1419
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Building Energy Consumption Prediction Using Grid Search-Optimized XGBoost: A Comparative Evaluation with Random Forest on Synthetic Data

Adinda Putri Pratiwi[1], Didi Susilo Budi Utomo[2*], Rihartanto[2]

[1] Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Samarinda, Indonesia
[2] Jurusan Teknologi Informasi, Politeknik Negeri Samarinda, Samarinda, Indonesia
Email: [1]app758@umkt.ac.id, [2]dsbudiutomo10@gmail.com, [3]rihartanto@gmail.com
Email Penulis Korespondensi: dsbudiutomo10@gmail.com*
Submitted **08-07-2026**; Accepted **23-10-2026**; Published **31-10-2026**

Abstract

The rising energy demands of buildings necessitate systems for efficient energy usage. Consequently, accurate prediction methods are required to assist in energy planning, control, and optimization. This study aims to predict energy consumption using the XGBoost algorithm, with parameter optimization performed via Grid Search. Additionally, Random Forest is employed as a benchmark and comparative method, and feature importance analysis is conducted to identify the variables that most significantly influence energy consumption. The dataset used is a synthetic dataset obtained from Kaggle. The dataset simulates environmental conditions and energy use in buildings. The results indicate that the best-performing model is the parameter-optimized XGBoost, which outperformed both Random Forest and the default XGBoost model. Specifically, the optimized XGBoost model achieved an MAE of 4.0029 ± 0.0127, an RMSE of 4.9352 ± 0.0104, and an R^2 of 0.6026 ± 0.0017. Feature importance analysis identified temperature as the most important feature for predicting energy consumption in the evaluated model. However, because the study is using synthetic data rather than real data from actual building measurements, the findings may not fully represent real-world buildings and should not be directly generalized without further validation. Further evaluation using real-world building energy consumption data is necessary to assess the practical applicability of the approach.

Keywords: Energy Consumption; Prediction; Machine Learning; XGBoost; Grid Search.

1. INTRODUCTION

The increasing demand for energy due to population growth, industry, and urbanization makes resource management a critical issue. Global energy demand is predicted to increase along with increased activity. Increased demand that is not balanced by proper management will create problems and adversely impact other sectors, such as the economy and the environment [1]. Therefore, an efficient energy management strategy is needed to support sustainable energy use in the future. Energy management is needed to ensure efficient use by predicting the amount of energy needed. Accurate predictions can aid planning, control, and optimization of energy use in various sectors.

Rapid technological developments have driven various studies on how to manage energy efficiently, one of which is the use of machine learning. Machine learning is able to learn patterns between variables, thus producing more accurate predictions [2]. Several previous studies have utilized machine learning for prediction. One study used SVM to predict electricity use in homes [3], but did not use parameter optimization and feature importance analysis. Another study used Random Forest to predict energy production from photovoltaics [4], but did not use parameter optimization. Furthermore, research using XGBoost and Random Forest to predict energy consumption in electric vehicles to optimize mileage [5] has not yet analyzed which variables are most influential on energy consumption. Furthermore, other research using XGBoost to predict energy consumption in steel mills has utilized feature importance analysis [6], but has not optimized the model parameters used. Based on previous research, a research gap inside the lack of model parameter optimization, particularly for XGBoost, while simultaneously analyzing the most influential variables in increasing energy consumption, and comparing other models as benchmarks within a single study. XGBoost was chosen because it provides good predictions in prediction cases [7], handles nonlinear relationships [8], and reduces overfitting [9].

Various studies have shown that machine learning, particularly XGBoost, can produce good performance. Parameter selection in machine learning significantly influences model performance in learning data patterns. Inappropriate parameter selection can lead to overfitting or underfitting of the model [10]. The best combination of parameters is needed to produce optimal evaluation. Therefore, hyperparameter tuning is needed. Hyperparameters are used in machine learning to generate the best parameter combination to improve model performance [11].

In addition to producing accurate predictions, interpreting the factors influencing energy consumption is also a crucial aspect of energy management. Feature importance analysis can be used to identify which variables most influence prediction results [12]. The results of feature importance analysis can be used as the basis for recommendations [13] for decision-making related to energy strategy and energy efficiency. However, research discussing the combination of hyperparameter and feature importance analysis is still limited. Therefore, research is needed to demonstrate opportunities to improve predictive performance while enhancing the interpretability of models that influence energy consumption.

Based on the existing description, this study proposes an approach to building an XGBoost model to predict energy consumption by optimizing parameters using Grid Search. Furthermore, this study uses Random Forest as a benchmark and provides a comparative evaluation of model performance, while XGBoost with default hyperparameters is also evaluated to determine the effect of hyperparameter optimization. Feature importance analysis is conducted to

Copyright © 2026 The **Adinda Putri Pratiwi**, Page 1412
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9980
Hal 1412-1419
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

identify the variables that contribute most to the prediction results. The proposed approach is based on the hypothesis that hyperparameter optimization using Grid Search can improve predictive performance of XGBoost compared with its default configuration. The novelty of this study lies in the integrated evaluation of predictive accuracy, performance consistency across ten experimental trials, computational cost, and feature importance within the same experimental framework. This evaluation provides a comparative assessment of the trade-offs between predictive performance, computational efficiency, and feature importance for the synthetic dataset used in this study. The findings are intended to characterize model behavior under the experimental conditions considered and should not be interpreted as direct evidence of performance or generalizability in real-world buildings. Further validation using measured building energy consumption data is needed to assess the practical applicability of the approach.

2. METHODOLOGY

This research uses several stages as shown in Figure 1. The first stage is collecting the electricity consumption dataset. The second stage is EDA (Exploratory Data Analysis), which is used to understand the data and detect potential problems in the data. The third stage is preprocessing. The stages carried out include checking for null data and converting time feature extraction to make it easier to recognize. The fourth stage is dividing the data into training and testing data. In the data division stage, the data is divided into 80% for training and 20% for testing using chronological split. In addition, TimeSeriesSplit with five splits was used for cross-validation during Grid Search hyperparameter optimization to preserve the temporal order of observations during model selection. The fifth stage introduces the Random Forest model, the default XGBoost, and the tuned XGBoost. The tuned XGBoost uses hyperparameters by optimizing the parameters n_estimators, max_depth, learning_rate, and subsample. The sixth stage is evaluation using MAE, RMSE, and R^2. The final stage is feature importance analysis.

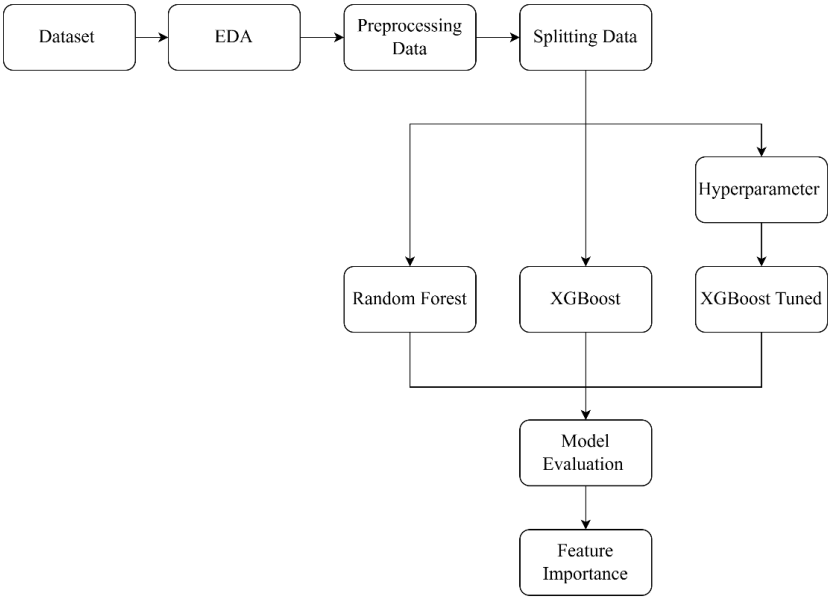


Figure 1. Research Framework

2.1 Dataset

The dataset is a public dataset obtained from Kaggle under the name Energy Consumption Prediction [14]. According to the dataset description, the dataset represents a hypothetical environment designed to emulate real-world energy consumption scenarios. The dataset is synthetic and it is designed to simulate building energy consumption based on 10 variables. The synthetic data used represents the relationship between variables that resemble real conditions without using real operational data. The dataset has variables such as timestamp, temperature, humidity, occupancy, HVAC, square footage, lighting usage, around-the-clock renewable energy, holidays, and energy consumption as research targets. The dataset has a total of 1000 data points.

2.2 Random Forest

Random Forest is an ensemble-based algorithm built on decision trees that works by randomly constructing a number of decision trees and combining the prediction results from all the trees to produce a final prediction [15]. In regression problems, the final prediction is obtained by calculating the average prediction results from all the trees formed. The more trees there are, the greater the accuracy of the prediction and reduce overfitting [16]. Therefore, Random Forest is widely used in various prediction problems, especially energy consumption. In this study, Random Forest is used as a benchmark model to compare performance with the proposed XGBoost algorithm whose parameters have been optimized. The

Copyright © 2026 The **Adinda Putri Pratiwi**, Page 1413
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9980
Hal 1412-1419
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Random Forest model uses the default hyperparameter configuration provided by scikit-learn, without additional hyperparameter optimization. The parameter that were used in Random Forest are max_depth None, min_samples_leaf 1, n_estimators 100, and max_features 1.0.

2.3 XGBoost

XGBoost is a Gradient Boosting Decision Tree (GBDT) algorithm widely used in machine learning for classification and prediction [17]. XGBoost works by incrementally building decision trees, with each new tree aiming to correct the prediction errors generated by the previous tree. The algorithm has several advantages in handling nonlinear relationships and mitigating the risk of overfitting [8].

In the baseline XGBoost model, the hyperparameter used were n-estimators 100, max_depth 6, learning rate 0.3, and subsample 1.0. The baseline model was trained using the training set and evaluated on the test set to establish the initial performance of XGBoost before hyperparameter optimization. Since the dataset consists of temporal energy data, a chronological data split was applied to preserve the temporal order, where earlier observations were used for training and subsequent observations were used for testing. Random shuffling was not applied during the train-test split to prevent information from future observations from being introduced into the training process.

To improve XGBoost performance, this study used the Grid Search method for parameter optimization. Grid Search is a systematic parameter search method that tests all combinations of predetermined parameter values [18]. Grid Search uses a cross-validation mechanism to perform the model parameter testing process without the need for manual validation one by one c[19]. Grid Search hyperparameter optimization was performed using five-fold time-series cross-validation (TimeSeriesSplit, with n_splits=5). The negative mean absolute error (neg_mean_absolute_error) was used as the scoring metric to select the best hyperparameter configuration. In this study, the optimized parameters include n_estimators, max_depth, learning rate, and subsample. The model with the best evaluation results from Grid Search will be used as the final model for energy prediction process. The parameters used in this study are shown in Table 1.

Table 1. Grid Search Parameter

Parameter	Value
n_estimators	100, 200, 300
max_depth	3, 5, 7
learning_rate	0.01, 0.05, 0.1
Subsample	0.8, 1.0

2.4 Model Evaluation

Evaluation was conducted to measure the performance of the model in predicting energy consumption. In this study, 3 model evaluation metrics were used: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2. MAE is used to calculate the average value of the deviation between the actual value and the predicted value [20]. MAE is used to describe how much prediction error is made by the model. RMSE calculates the average squared error value in prediction between the actual value and the predicted value; the smaller the value, the better the model. RMSE is used to determine the model's ability to minimize extreme prediction errors, so it can show the model's sensitivity [21]. R^2 is a score to see how much variation in the dependent variable can be explained by the model [22]. So, the greater the R^2 value, the more variation can be explained by the model.

To ensure a comprehensive assessment, the model was evaluated across ten trials using different random seeds generated by NumPy's random number generator initialized with a fixed seed of 42. Ten distinct seeds were randomly choosen from 1 to 99 without replacement. Each seed is selected to control the algorithm, and the data training and testing remain fixed for all trials. These varied experiments allow for the observation of performance fluctuations across repeated runs. The mean and standard deviation were calculated from these ten trials to determine the average prediction performance and the extent of result variation; the mean represents the average value across all trials, while the standard deviation indicates the variability in model performance.

3. RESULT AND DISCUSSION

3.1 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a step to understand the characteristics of a dataset before the capitalization process. The purpose of EDA is to identify relationships between variables. This study used a correlation heatmap to demonstrate the level of relationship between variables in a dataset. The correlation heatmap is shown in Figure 2.

Based on the correlation analysis, the temperature variable has the highest correlation value with energy consumption at 0.70, indicating a positive relationship; the higher the temperature, the higher the energy consumption. The HVAC (Heating, Ventilation, and Air Conditioning) variable, or room cooling system, has a value of 0.29, indicating a weak-to-moderate positive relationship, meaning the use of air conditioning systems will contribute to increased energy consumption. Therefore, higher outdoor temperatures and the use of air conditioning will increase energy consumption. Third, the variables of occupancy, lighting, and renewable energy have values of 0.19, 0.093, and 0.081, respectively, indicating a weak positive correlation, meaning they contribute little to increased energy consumption.

Copyright © 2026 The **Adinda Putri Pratiwi**, Page 1414
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9980
Hal 1412-1419
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

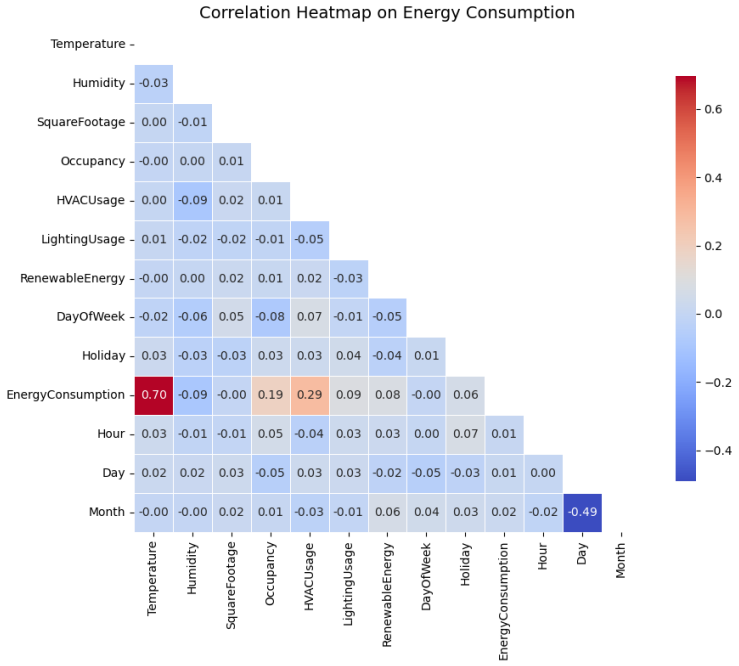


Figure 2. Correlation Heatmap

The humidity variable has a weak negative correlation of -0.094, indicating a weak negative linear association between the two variables. Square footage has a correlation coefficient close to zero at -0.0011, indicating a negligible linear association with energy consumption in the evaluated dataset. Similarly, the correlation analysis indicates a weak linear association between the time-related variables and energy consumption. However, near-zero correlation coefficients do not necessarily imply that these variables have no predictive relevance, as nonlinear relationships or interactions with other variables may exist. The EDA results indicate that all variables can be used in the modeling process.

3.2 Discussion

In this study, the modeling process was carried out using three approaches: Random Forest, which was used as a benchmark model, XGBoost default via its default parameter, and XGBoost tuned were used in hyperparameter optimization setting using Grid Search. To maintain the consistency of all model prediction, the data for training and evaluation had the same partitions for the sake of comparison. All model was evaluated through ten experimental trials with different seeds to obtain more reliable comparisons with a fixed random number from the generator seed, enabling the experiment to be reproduced under same configuration.The same training and testing partitions were maintained across all trials, while the model random states varied according to the predefined seed list. The evaluation of the Random Forest, XGBoost default, and XGBoost tuned models is presented in Table 2. The evaluation was carried out using the MAE, RMSE, and R^2 metrics. MAE and RMSE are used to measure the level of prediction error, while the R^2 value is used to determine the model's ability to explain data variations. Models with small MAE and RMSE values and large R^2 are models that have good performance. The mean and standard deviation of the evaluation metrics were used to assess both predictive performance and consistency across trials.

Table 2 shows that the tuned XGBoost performed best compared to other algorithms. This is evident from the MAE value of 4.0029 ± 0.0127 and RMSE of 4.9352 ± 0.0104, which are the lowest compared to other models. Furthermore, the R^2 value of 0.6026 ± 0.0017 indicates that the tuned XGBoost can better explain data variation. However, the Random Forest has a slightly lower than the tuned XGBoost. Random Forest achieved an MAE of 4.2067 ± 0.0487, an RMSE of 5.1559 ± 0.0482, and an R^2 of 0.5662 ± 0.0081. The difference in R^2 between Random Forest and tuned XGBoost was only 0.0364, suggesting that their predictive performance was similiar in the terms of proportion on variance.

Table 2. Result of Modeling

Model	MAE(Mean±SD)	RMSE(Mean±SD)	R^2(Mean±SD)	Time(Mean±SD)
Random Forest	4.2067 ± 0.0487	5.1559 ± 0.0482	0.5662 ± 0.0081	0.4971 ± 0.1286
XGBoost Default	4.5071 ± 0.0000	5.4524 ± 0.0000	0.5149 ± 0.0000	0.2406 ± 0.0663

Copyright © 2026 The **Adinda Putri Pratiwi**, Page 1415
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9980
Hal 1412-1419
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

XGBoost Tuned	4.0029 ± 0.0127	4.9352 ± 0.0104	0.6026 ± 0.0017	16.5962 ± 3.4831

The results for the default XGBoost model were an MAE of 4.5071 ± 0.0000, an RMSE of 5.4524 ± 0.0000, and an R^2 of 0.5149 ± 0.0000; an improvement of 0,0877 was observed, along with a reduction in the MAE and RMSE values. The results indicate that using Grid Search for parameter optimization improves evaluation results compared to using the default model parameters. Based on Table 2, it shows that the hyperparameter configuration that is selected by Grid Search produced more accurate predictions than the default model to evaluate the dataset. The Grid Search were examined through 54 hyperparameter combinations.

The mean value represents the overall predictive performance, while standard deviation provides an indication if the consistency across different experimental conditions. A lower standard deviation indicates that the model produced more consistent results across the trials, whereas a higher standard deviation indicates greater variation in model performance. The results of the mean and standard deviation are shown in Table 2 The tuned XGBoost had a lower standard than other model, meaning it has smaller variation. It is shown too in Figure 3, where the tuned XGBoost is more consistence than the other model. Figure 3 Shows the variation across ten trials by showing MAE, RMSE, and R^2. It shows that XGBoost default had higher error values than the other models, while tuned XGBoost and Random Forest values remained relatively close across the ten trials.

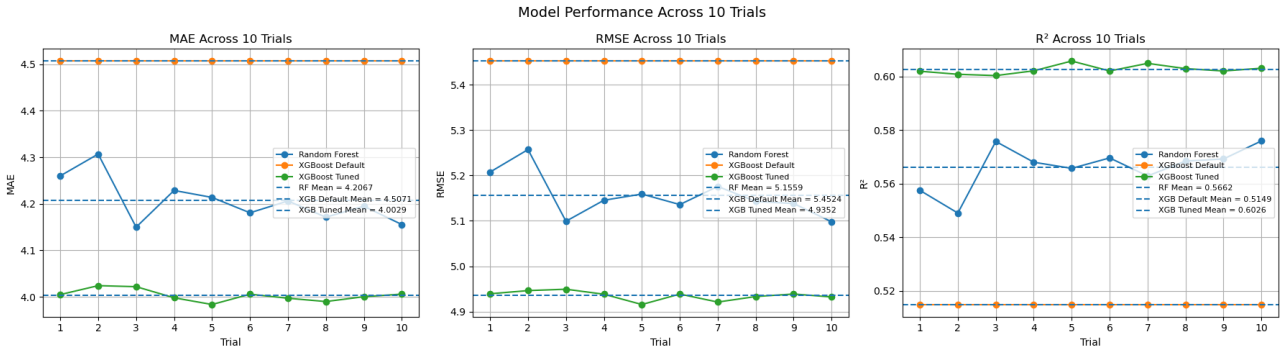


Figure 3. Model Performance Across 10 Trials

The parameter combination obtained in the tuned XGBoost is learning rate: 0.01, max_depth: 3, n_estimator: 300, and subsample: 0.8. This combination of parameters produces a balanced model, thereby improving prediction accuracy. Performance improvements in the tuned XGBoost are influenced by the right combination of parameters. In this experimental setting, a lower learning rate may improve optimization stability by allowing the model to update its predictions more gradually during training. Using a max_depth of 3 results in a simpler decision tree. This combination of parameters makes the model more stable, resulting in better generalization capabilities. However, their effects depend on the dataset and experimental configuration and do not necessarily guarantee better generalization in other settings.

However, in XGBoost Tuned also, the time execution took longer because it needs additional computational resources to process the Grid Search procedure, which evaluates 54 hyperparameter combinations through time-series cross-validation. The reported execution time includes the cross-validation process and the subsequent refitting of the best estimator, where the test-set prediction and evaluation metric calculation are excluded from the measurement. The potential predictive performance should be considered alongside the additional computational cost when assessing the suitability of the optimized model. Given the small difference in predictive performance between tuned XGBoost and Random Forest, model selection should consider both predictive accuracy and computational requirements. Random Forest may be a practical alternative when computational efficiency is a priority, while tuned XGBoost may be preferred when achieving the lowest observed prediction errors is the primary objective.

To complement the aggregate performance metrics, Figure 4 presents the actual versus predicted energy consumption values and residual plots for the three models in Trial 4. In the scatterplot, the horizontal axis represents the actual values, while the vertical axis represents the predicted values. The diagonal line represents the ideal distance between predicted and actual values. The closer the points are to the diagonal line, the smaller the model's prediction error. Figure 4 shows that the model has a distribution pattern that follows the diagonal line, indicating that the model is capable of producing predictions with relatively small errors. In the default XGBoost, the points are still scattered away from the diagonal line, indicating that there are some prediction errors in some observations. This indicates that, in its predictions, the default XGBoost still produces larger prediction errors compared to the two other models.

Copyright © 2026 The **Adinda Putri Pratiwi**, Page 1416
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9980
Hal 1412-1419
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

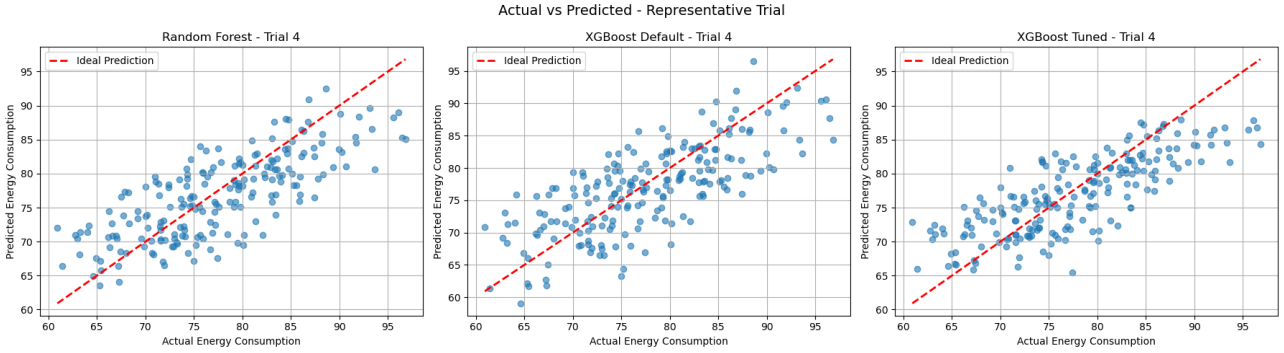


Figure 4. Scatter Plot Actual Vs Predicted

In both Random Forest and XGBoost, the points are more centrally located, indicating that the predicted values are closer to the actual values, thus improving accuracy compared to the default XGBoost model. This is consistent with the results shown in the smaller MAE and RMSE values, as well as the higher R^2. In the tuned XGBoost model, the points are also closer to the diagonal line, indicating that parameter optimization has successfully improved the model's ability to predict energy consumption. Moreover, since Figure 4 presents only Trial 4, the scatter plots serve as an illustrative example of model predictions rather than a complete assessment of performance across all ten trials.

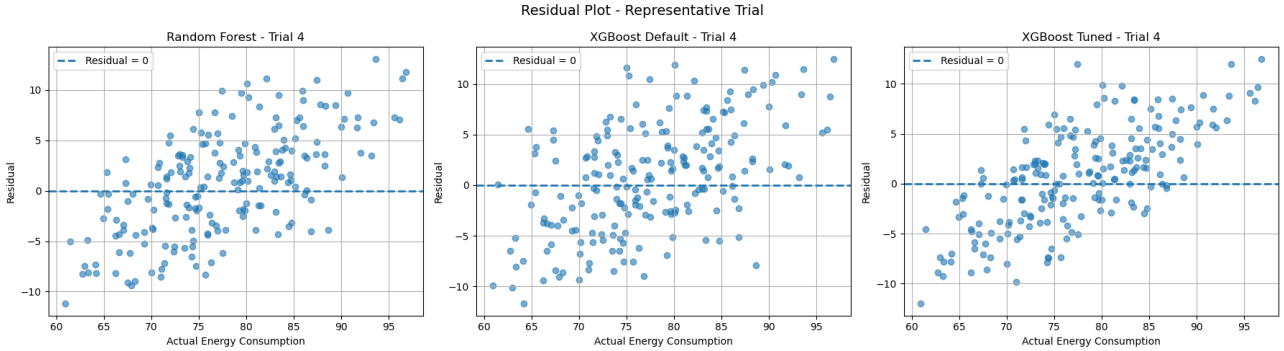


Figure 5. Residual Plot

Residual plots from Random Forest, the default XGBoost, and the tuned XGBoost are shown in Figure 5. Residual plots are used to evaluate the prediction error generated by the model by comparing the residual values to the predicted values. According to Figure 5, the residuals are randomly distributed around the zero line without forming any specific pattern, indicating that the prediction error is random and there is no indication of systematic bias in the model.

In Random Forest and the tuned XGBoost, the points have a narrow range, indicating the model is able to predict energy consumption well. No specific patterns are found in the plot indicating the model is experiencing systematic bias in its predictions. Meanwhile, with the default XGBoost, the points are more spread out, indicating that the model does not predict as well as the previous two models, resulting in a larger prediction error. The wider spread of residuals for default XGBoost is consistent with its higher MAE and RMSE in Table 2. Because Figure 5 shows residuals from only one trial, the observed patterns should be interpreted together with the ten-trial results in Table 2 and Figure 3. This combined evaluation provides a more balanced interpretation of predictive accuracy and observed consistency

3.3 Feature Importance Analysis

To determine the variables that most influence energy consumption, a feature importance analysis was conducted. The results of the feature importance analysis can be seen in Figure 6. The Feature importance analysis provides insight into the relative predictive contributions of the input features and their potential relationships in building energy consumption prediction. Feature importance was measured using the Gain metric, and the raw scores were normalized so that the sum of the importance values across all input features equaled as one. Temperature is the feature that most influences the increase and decrease in energy consumption, because it has the highest normalized importance value compared to other variables. The feature importance value for the temperature variable is 41.68% from the total importance. This indicates that temperature contributed more to the model's predictive performance than the other evaluated features. Higher temperatures may increase cooling demand, leading to greater HVAC system usage and, consequently, higher building energy consumption. This potential relationship may help explain the high importance assigned to temperature by the model. Holiday was the feature with the lowest feature importance, with a value of 2.85%. This indicates that holidays do not have a significant effect on predicting energy consumption. However, these scores represent predictive relevance within the fitted model and do not establish causal relationships with energy consumption.

Copyright © 2026 The **Adinda Putri Pratiwi**, Page 1417
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9980
Hal 1412-1419
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

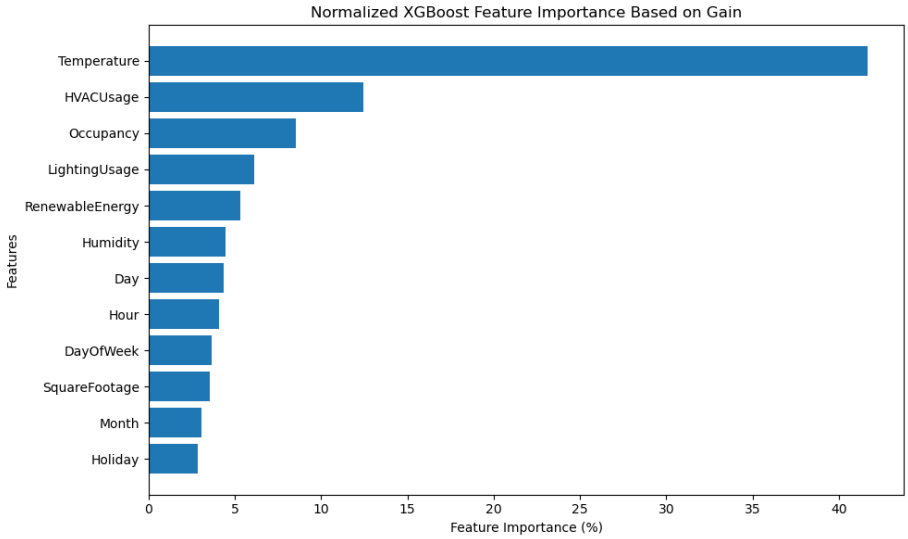


Figure 6. Analysis Feature Importance

4. CONCLUSION

Based on the results of the research, it can be concluded that the XGBoost algorithm, with optimized parameters, performs best in predicting building energy consumption compared to Random Forest and the default XGBoost. The results of the tuned XGBoost are indicated by lower MAE and RMSE values and higher R^2 values. The most influential variables are temperature, HVAC, and occupancy, which are the variables with the highest importance values. The variables with the least influence on energy consumption are holidays, square footage, and humidity. These findings provide insights into the variables associated with predicted building energy consumption in the dataset and may help guide further investigations into building energy management and efficiency strategies. However, there is limitation in this study, the use of a synthetic dataset generated to simulate building energy consumption rather than data collected from actual buildings. Although synthetic data provide a controlled environment for evaluating machine learning models, the patterns represented in the dataset may not fully capture the complexity and variability of real-world data. Further validation using real building energy consumption data is therefore necessary to assess the practical applicability of the proposed approach. Future research can use real-world datasets, incorporate feature engineering techniques, and compare XGBoost with other algorithms such as LightGBM, CatBoost, and Deep Learning-based models to investigate whether these approaches can improve predictive performance

REFERENCES

[1] R. H. Melo, F. Moko, and S. E. Saleh, "Tantangan Pembangunan Sumberdaya Alam di Indonesia: Dampak Lingkungan dan Ekonomi dalam Pencapaian Keberlanjutan," *J. Penelit. Geogr.*, vol. 3, no. 2, pp. 149–154, 2024, doi: 10.37905/geojpg.v3i2.29544.

[2] P. Triya, N. Suarna, and N. Dienwati Nuris, "Penerapan Machine Learning Dalam Melakukan Prediksi Harga Saham Pt. Bank Mandiri (Persero) Tbk Dengan Algoritma Linear Regression," *J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 1207–1214, 2024, doi: 10.36040/jati.v8i1.8958.

[3] N. S. Syam et al, "Model Support Vector Machine untuk Prediksi pada Penggunaan Energi Listrik di Rumah Hemat Energi," *J. Inform.*, vol. 1, no. 2, pp. 56–59, 2022, doi: 10.57094/ji.v1i2.360.

[4] E. P. Febtiawan, L. A. S. I. Akbar, and A. S. Rachman, "Forecasting Produksi Energi Photovoltaic Menggunakan Algoritma Random Forest Classification," *J. Inf. Syst. Res.*, vol. 5, no. 4, pp. 1053–1062, 2024, doi: 10.47065/josh.v5i4.5514.

[5] M. R. R. Saelan, R. L. Hasanah, and S. Fauziah, "Optimalisasi Prediksi Jarak Tempuh Kendaraan Listrik Menggunakan Xgboost Dan Feature Selection Random Forest," *INTI Nusa Mandiri*, vol. 20, no. 2, pp. 213–221, 2026, doi: 10.33480/zahvvq98.

[6] H. Utama and J. D. Santoso, "Pemodelan prediktif konsumsi energi listrik di pabrik baja berbasis XGBoost untuk pengelolaan sumber daya berkelanjutan," *J. Multidisiplin Bahtera*, vol. 2, no. 3, pp. 109–118, 2025, doi: 10.24432/C52G8C.

[7] D. T. Murdiansyah, "Prediksi Stroke Menggunakan Extreme Gradient Boosting," *J. Inform. dan Komput.*, vol. 8, no. 2, p. 419, 2024, doi: 10.26798/jiko.v8i2.1295.

[8] I. Maulita and A. M. Wahid, "Prediksi Magnitudo Gempa Menggunakan Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron Berdasarkan Data Kedalaman dan Geolokasi," *J. Pendidik. dan Teknol. Indones.*, vol. 4, no. 5, pp. 221–232, 2024, doi: 10.52436/1.jpti.470.

[9] M. B. Prayogi, F. Apriani, and Nirma, "Prediksi Angka Harapan Hidup Menggunakan Random Forest dan XGBoost Regression," *J. Inform. dan Teknol.*, vol. 2, no. 1, pp. 112–121, 2025, doi: 10.30599/ht1k1h53.

[10] Putri. S. K. N. A et al., "Penerapan Hyperparameter Tuning pada Model Klasifikasi untuk Prediksi Risiko Penyakit Jantung : Implementation of Hyperparameter Tuning for Classification Models in Heart Disease Risk Prediction," *MALCOM Indones.*

Copyright © 2026 The **Adinda Putri Pratiwi**, Page 1418
This Journal is licensed under a Creative Commons Attribution 4.0 International License

J. Mach. Learn. Comput. Sci., no. Vol. 5 No. 4 (2025), pp. 1181–1189, 2025, doi: 10.57152/malcom.v5i4.2138.

[11] M. Y. Shams, A. M. Elshewey, E. S. M. El-kenawy, A. Ibrahim, F. M. Talaat, and Z. Tarek, "Water quality prediction using machine learning models based on grid search method," *Multimed. Tools Appl.*, vol. 83, no. 12, pp. 35307–35334, 2024, doi: 10.1007/s11042-023-16737-4.

[12] M. H. K. Anam, D. Kurnianingtyas, and A. A. Soebroto, "Implementasi Algoritma Random Forest Untuk Prediksi Churn Pada Pelanggan Retail Online," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 10, no. 4, pp. 1–10, 2026.

[13] L. N. Hapsari, D. Ridhwanullah, P. Studi, S. Informasi, and U. T. Serangkai, "Prediksi Kelulusan Siswa Sekolah Menengah Menggunakan XGBoost dan Analisis Feature Importance," vol. 15, pp. 0–6, 2026.

[14] Mr.Simple, "Energy-consumption-prediction," Kaggle. Accessed: Jun. 09, 2026. [Online]. Available: https://www.kaggle.com/datasets/mrsimple07/energy-consumption-prediction/data

[15] L. Sari, A. Romadloni, and R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Infotekmesin*, vol. 14, no. 1, pp. 155–162, 2023, doi: 10.35970/infotekmesin.v14i1.1751.

[16] H. Rathore, H. K. Meena, and P. Jain, "Prediction of EV Energy consumption Using Random Forest And XGBoost," *Proc. - 2nd Int. Conf. Power Electron. Energy, ICPEE 2023*, pp. 1–6, 2023, doi: 10.1109/ICPEE54198.2023.10060798.

[17] H. Abbasimehr, R. Paki, and A. Bahrini, "A novel XGBoost-based featurization approach to forecast renewable energy consumption with deep learning models," *Sustain. Comput. Informatics Syst.*, vol. 38, no. November 2022, p. 100863, 2023, doi: 10.1016/j.suscom.2023.100863.

[18] S. L. Sari and B. Rahmat, "Penerapan Algoritma Grid Search Untuk Memprediksi Customer Churn Pada Bank Menggunakan Perbandingan Optimasi Decision Tree Dan Random Forest," *J. Mhs. Tek. Inform.*, vol. 10, no. 1, pp. 908–913, 2026, doi: 10.36040/jati.v10i1.16819.

[19] S. W. Irianandа, R. W. Budiawan, A. Y. Rahman, and I. Istiadi, "Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan Svm, Grid Search Dan Kombinasi N-Gram," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 4, pp. 743–752, 2024, doi: 10.25126/jtiik.1148244.

[20] I. J. Matondang, E. Budianita, F. Syafria, and I. Afrianty, "Sistem Prediksi Produksi Kelapa Sawit Berbasis Gradio Menggunakan Algoritma Regresi Linear Berganda," *Bull. Comput. Sci. Res.*, vol. 6, no. 2, pp. 664–672, 2026, doi: 10.47065/bulletincsr.v6i2.994.

[21] M. R. R. Saelan, R. L. Hasanah, and S. Fauziah, "Optimalisasi Prediksi Jarak Tempuh Kendaraan Listrik Menggunakan Xgboost Dan Feature Selection Random Forest," *LPPM Nusa Mandiri*, vol. 19, no. 2, pp. 146–153, 2025, doi: inti.v20i2.7828.

[22] J. Gao, "R-Squared (R2):-How much variation is explained?," *Res. Methods Med. Heal. Sci.*, vol. 5, no. 4, pp. 104–109, 2024, doi: 10.1177/26320843231186398.

Copyright © 2026 The **Adinda Putri Pratiwi**, Page 1419
This Journal is licensed under a Creative Commons Attribution 4.0 International License