Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 8, No. 1, March 2025, pp. 76 – 85
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Performance Evaluation of Machine Learning Models for Predicting Household Energy Consumption: A Comparative Study

Gregorius Airlangga
Information System Study Program, Atma Jaya Catholic University of Indonesia, Indonesia
Email: gregorius.airlangga@atmajaya.ac.id

Article Info

Article history:
Received Sep 24th, 2024
Revised Nov 29th, 2024
Accepted Dec 5th, 2024

Keyword:
Energy Consumption Prediction
Ensemble Methods
Machine Learning
Regression Models
Time Series Forecasting

ABSTRACT

Accurate prediction of household energy consumption is critical for improving energy efficiency and optimizing resource allocation in smart grids. This study evaluates the performance of several machine learning regression models, including Linear Regression, Ridge Regression, Lasso Regression, Random Forest, Gradient Boosting, XGBoost, CatBoost, and LightGBM, for predicting daily household energy consumption. The models were trained and tested on time series data, and their performance was measured using four key metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2. Results show that non-linear models, especially ensemble-based methods such as Random Forest and CatBoost, outperformed traditional linear regression models. Random Forest achieved the lowest MAE (0.1682) and competitive RMSE (0.2450), making it the best overall model. CatBoost, with its advanced gradient boosting algorithm, also demonstrated superior predictive accuracy, achieving an RMSE of 0.2421 and an MAE of 0.1830. In contrast, linear models struggled to capture the complex patterns in the data, with Linear Regression showing the worst performance. The negative R^2 scores across all models indicate challenges in explaining the variance in the dataset, which may be attributed to external factors or noise not captured by the models. This study highlights the importance of choosing appropriate machine learning models for time series forecasting and recommends further exploration of deep learning models and external features to improve prediction accuracy.

Corresponding Author:
Gregorius Airlngga,
Information System Study Program, Atma Jaya Catholic University of Indonesia
Jakarta, Indonesia
Email: gregorius.airlangga@atmajaya.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v8i1.32791

1. INTRODUCTION

The increasing demand for efficient energy management in residential buildings has become a significant concern in the face of rising energy costs and the need for sustainability [1]–[3]. As households integrate renewable energy sources such as solar power, and smart home devices become more prevalent, predicting energy consumption accurately has emerged as a critical challenge [2], [4], [5]. Efficient forecasting of energy demand enables better energy usage, cost reductions, and improved integration of renewable sources [6]. In this context, machine learning models, particularly regression-based models, offer promising solutions for forecasting time series data, which is crucial for optimizing residential energy systems [7]. Time series forecasting plays a vital role in predicting energy consumption and generation, especially for managing residential energy needs [8]. Accurate forecasting helps homeowners optimize energy use, informs grid operators about real-time demand fluctuations, and supports energy suppliers in balancing supply and demand [9]–[11]. Energy consumption is highly dynamic, influenced by factors such as weather, time of day, and appliance usage [12]. The challenge lies in capturing these complex patterns and dependencies, which requires advanced machine learning techniques capable of modeling nonlinear relationships [13].

As renewable energy sources, particularly solar energy, are becoming more integrated into residential power systems, precise forecasting is essential [14]. This is because renewable energy is often intermittent, making it challenging to predict both generation and consumption patterns [15]. Additionally, smart meters and Internet of Things (IoT) devices continuously generate large amounts of data, which can be leveraged by advanced machine learning models to improve energy forecasting accuracy [16]. In the field of energy consumption forecasting, early models were dominated by statistical methods such as Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing [17]. These models are favored for their simplicity and ability to explain linear relationships. However, with the growing complexity and nonlinearity in energy data, machine learning models have become increasingly popular.

Linear regression remains a foundational approach due to its simplicity and interpretability, yet it struggles with capturing nonlinearity in the data [17]. To address this, models like Ridge and Lasso regression have been introduced, which incorporate regularization techniques to improve model generalization and prevent overfitting [18]. However, these models still fall short in accurately predicting complex, nonlinear energy consumption patterns. Ensemble models such as Random Forest and Gradient Boosting have shown remarkable improvements in predictive performance [19]. Random Forest, by aggregating multiple decision trees, enhances accuracy and reduces overfitting. Gradient Boosting, through sequential tree-building, excels at refining predictions by minimizing errors from previous models. Both approaches handle nonlinearity effectively and have been widely applied to energy forecasting problems [20].

More advanced techniques, such as XGBoost, CatBoost, and LightGBM, represent the latest developments in machine learning for regression tasks. XGBoost has gained widespread adoption due to its scalability and performance in handling large datasets [21]. CatBoost, which is designed to handle categorical data efficiently, and LightGBM, known for its fast training and high accuracy, have further extended the capabilities of gradient boosting methods in time series forecasting [22]. The global shift toward renewable energy sources amplifies the need for precise energy demand and generation forecasting [23]. As residential buildings increasingly rely on solar panels and other intermittent energy sources, accurate forecasting helps ensure energy reliability and stability [24]. Moreover, balancing energy production with consumption is essential for both consumers and suppliers to minimize costs, reduce waste, and maintain grid stability.

Additionally, with the rise of smart homes and IoT devices, there is an ever-growing stream of real-time energy consumption data. This provides a significant opportunity for developing machine learning models that can process and analyze this data for predictive insights. Failure to harness these advancements could lead to inefficiencies and increased costs in energy management [25]. Currently, the state of the art in energy forecasting blends traditional statistical approaches with advanced machine learning techniques. While models like ARIMA are still used in some cases, they are increasingly being replaced by machine learning methods, which are better suited to capturing the nonlinear, complex nature of energy consumption data [26].

Among the top-performing models, Random Forest and Gradient Boosting have shown excellent results in a variety of forecasting tasks. These models excel at handling the intricate interactions between various features influencing energy consumption, such as weather conditions, appliance usage, and temporal patterns [27]. Recent innovations like XGBoost, CatBoost, and LightGBM have further improved accuracy and computational efficiency, making them the preferred choice in many energy prediction studies. Despite their success, these models present challenges in terms of interpretability and computational demands. While their complexity allows for better performance, it also makes them harder to interpret compared to simpler linear models [28]. This trade-off between performance and transparency is an ongoing challenge in energy forecasting.

The objective of this study is to evaluate and compare various machine learning regression models in predicting household energy consumption using time series data. The models under consideration include traditional linear approaches such as Linear Regression, Ridge, and Lasso, as well as modern ensemble and gradient boosting techniques like Random Forest, Gradient Boosting, XGBoost, CatBoost, and LightGBM. The aim is to determine the most effective model for accurately forecasting energy usage in a residential setting, considering factors such as accuracy, computational efficiency, and interpretability. While significant research has been conducted on energy consumption forecasting, several gaps remain. First, there is limited research comparing a broad spectrum of regression models on residential energy data that includes external variables such as weather conditions. Most existing studies focus on either traditional statistical models or specific machine learning techniques, but few have comprehensively evaluated the latest ensemble and gradient boosting models in a comparative framework. Furthermore, most studies do not explore the use of rolling window sequences in time series forecasting, which can capture longer-term dependencies and improve prediction accuracy. Another gap lies in the trade-off between model performance and interpretability, with complex models often being more accurate but less transparent. This study aims to address these gaps by providing a detailed comparison of models, evaluating their performance and computational requirements, and offering insights into the trade-offs involved. The rest of the article is structured as follows: The next section

describes the dataset used in this study, including details on feature selection, data preprocessing, and scaling. This is followed by an explanation of the experimental setup, covering the models used, the evaluation metrics, and the training process. The results section presents a comparative analysis of the models' performance, highlighting their strengths and limitations in terms of accuracy and computational cost. Finally, the discussion and conclusion sections reflect on the findings, their implications for energy forecasting, and directions for future research.

2. RESEARCH METHOD

The methodology of this research outlines the design and procedure used to evaluate machine learning regression models for time series forecasting of household energy consumption. As presented in the figure 1, this process encompasses data acquisition, preprocessing, model selection, and evaluation, with a focus on ensuring that each step is scientifically sound and replicable. The objective is to compare the performance of various models and identify the most effective one for predicting household energy consumption. The method is described in a chronological order to reflect the research flow, supported by established techniques in time series forecasting and machine learning.

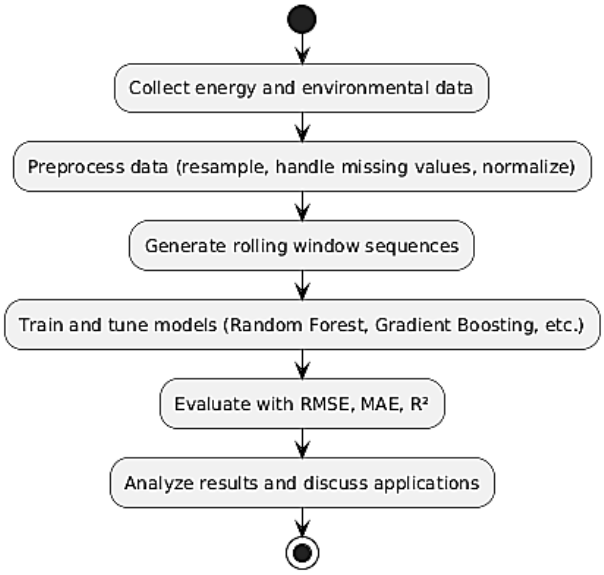

Figure 1. Research Methodology Process

2.1 Research Design

The research design revolves around predicting household energy consumption using machine learning regression models. The dataset contains time series data, and the rolling window technique is employed to capture temporal dependencies. In this setup, the model uses past data points (energy consumption over the last 30 days) to predict the next day's energy consumption. The models explored include traditional linear regression methods, along with more advanced ensemble and gradient boosting techniques.This study focuses on assessing the performance of several regression models by comparing their accuracy and efficiency in forecasting daily energy consumption. The key objective is to identify which model provides the most accurate predictions and the best trade-off between accuracy and computational efficiency.

2.2 Data Acquisition

The dataset used in this study was obtained from a smart home energy monitoring system, which recorded various parameters related to energy consumption and environmental conditions and can be downloaded from [29]. The primary features include total energy consumption, solar energy generation, and net power usage within the household. In addition to energy metrics, environmental variables such as temperature, humidity, wind speed, and cloud cover were recorded. The dataset spans several months, with observations made at hourly intervals, capturing both energy consumption patterns and external factors influencing energy usage. These variables are crucial for building robust machine learning models, as external factors like weather conditions can significantly affect energy consumption. The dataset contains continuous time series data, which is resampled to daily averages for this research. Resampling reduces short-term fluctuations and smooths out the inherent noise in hourly data, allowing the models to focus on long-term

consumption trends. Aggregating the data into daily intervals helps capture the overall energy consumption behavior without being overly influenced by transient appliance usage. The combination of energy usage data with weather information provides a more comprehensive dataset for building accurate predictive models.

2.3 Data Preprocessing

To prepare the data for model training, several preprocessing steps were performed. First, the raw data, initially recorded on an hourly basis, was resampled into daily averages to focus on daily consumption trends. The dataset was examined for missing values, as incomplete records can compromise the accuracy of machine learning models. Missing data points were imputed using linear interpolation. This method was selected because it preserves the temporal sequence by estimating values based on the trend of adjacent data points, ensuring minimal disruption to the time series' continuity. For categorical variables, a simple mode imputation was employed to replace missing values with the most frequent category. Furthermore, in order to tackle outliers problem, data were identified using the Interquartile Range (IQR) method to ensure that the dataset's integrity was maintained. Any data point lying beyond 1.5 times the IQR from the first and third quartiles was flagged as an outlier. These outliers were capped at the nearest valid data boundary to mitigate their potential impact without entirely removing them, which could distort temporal dependencies.

The resampling process reduces noise in the data caused by sporadic short-term appliance usage. If the hourly energy consumption data is denoted as x_i, then the daily average $\overline{x_d}$ for day d is computed as $\overline{x_d} = \frac{1}{n_d}\sum_{i=1}^{n_d} x_i$, where n_d is the number of hours in day d. Next, the data was scaled using Min-Max normalization. This ensures that all features are within a common range, typically $[0,1]$, which helps models converge more efficiently during training. For a given feature x, the scaled value x_{scaled} is calculated as follows $x_{\text{scaled}} = \frac{x-\min(x)}{\max(x)-\min(x)}$

Scaling is critical for models like gradient boosting, as it ensures that no feature dominates the training process due to differences in magnitude. To capture the temporal structure of the data, a rolling window approach was applied. This method creates sequences of past energy consumption values to serve as inputs to the model. The window size k was set to 30 days, meaning that the model predicts energy consumption on day t using the consumption values from days $t-30$ to $t-1$. Mathematically, the input sequence for time t, denoted as X_t, is represented as $X_t = [y_{t-30}, y_{t-29}, \ldots, y_{t-1}]$ where y_t is the target energy consumption on day t. The pseudocode for generating these rolling window sequences can be described as presented can be described as follows.

```
function create_sequences(data, time_steps):
    sequences = []
    labels = []
    for i = time_steps to len(data):
        sequence = data[i-time_steps:i]
        label = data[i]
        append sequence to sequences
        append label to labels
    return sequences, labels
\end{verbatim}
```

This process prepares the data for time series prediction by ensuring that the model has access to a sequence of historical data when making predictions.

2.4 Model Selection and Training

Several machine learning models were selected for this study, ranging from traditional linear regression to more advanced ensemble methods. Each model was trained on the preprocessed data, with the goal of predicting daily energy consumption based on the previous 30 days of consumption data. The first model considered was \textit{Linear Regression}, which assumes a linear relationship between the input variables and the target variable. The model can be expressed mathematically as $y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i +$ where β_0 is the intercept, β_i are the coefficients of the input features, X_i are the features themselves, and ϵ is the error term. This model serves as a baseline for comparing the performance of more complex models.

Next, Ridge Regression and Lasso Regression were applied, which are regularized versions of linear regression. Ridge regression introduces an L_2 regularization term that penalizes large coefficient values as presented as Ridge Loss $= \sum_{i=1}^{n}(y_i - \hat{y_i})^2 + \lambda \sum_{i=1}^{n} \beta_i^2$ where λ is a hyperparameter controlling the strength of the regularization. In contrast, Lasso regression introduces an L_1 penalty that encourages sparsity in the coefficients as presented as Lasso Loss $= \sum_{i=1}^{n}(y_i - \hat{y_i})^2 + \lambda \sum_{i=1}^{n} |\beta_i|$, both Ridge and Lasso help prevent overfitting by constraining the size of the model's coefficients.

The Random Forest Regressor was used as an example of an ensemble method. Random Forest constructs multiple decision trees during training, and the final prediction is the average of the predictions from all the trees $\hat{y} = \frac{1}{T}\sum_{j=1}^{T} f_j(X)$ where $f_j(X)$ is the prediction from the j-th decision tree, and T is the total number of trees. Then, the Gradient Boosting Regressor builds models sequentially, with each new model correcting the errors of the previous one. The final model is a weighted sum of the individual models as presented as $\hat{y} = \sum_{j=1}^{T} \alpha_j f_j(X)$ where α_j is the weight of the j-th model, and $f_j(X)$ is the prediction from the j-th model. More advanced gradient boosting models such as XGBoost, CatBoost, and LightGBM were also considered. These models enhance traditional gradient boosting by incorporating optimizations such as regularization, efficient handling of categorical features, and faster training with large datasets. Each model was trained on the sequences generated from the rolling window approach. The training process involved splitting the dataset into a training set (70%) and a test set (30%). The models were trained on the training data and evaluated on the test data.

2.5 Testing and Evaluations

Once the models were trained, their performance was evaluated using several standard metrics to measure the accuracy of the predictions. These metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2. The Mean Squared Error (MSE) measures the average squared difference between the predicted and actual values. It is given by $\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ where y_i represents the actual value, $\hat{y}_i$ is the predicted value, and n is the total number of observations. The Root Mean Squared Error (RMSE) is the square root of the MSE, providing an interpretable metric in the same unit as the target variable $\text{RMSE} = \sqrt{\text{MSE}}$, The Mean Absolute Error (MAE) measures the average absolute difference between the predicted and actual values $\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$. Finally, the R^2 score quantifies how well the model explains the variance in the target variable $R\text{^}2 = 1 - \square\{\sum\_\{i = 1\}\text{^}\{n\}\ (y\_i - \text{^}\{y\}\_i)\text{^}2\}\{\sum_{i=1}^{n}(y_i - \bar{y})^2$, where $\bar{y}$ is the mean of the actual values. The R^2 score ranges from 0 to 1, with higher values indicating better performance. The pseudocode for training and evaluating the models can be described as follows.

```
function train_and_evaluate(models, X_train, y_train, X_test, y_test):
    results = {}
    for model_name, model in models:
        model.fit(X_train, y_train)
        predictions = model.predict(X_test)
        predictions_rescaled = inverse_transform(predictions)
        y_test_rescaled = inverse_transform(y_test)

        mse = mean_squared_error(y_test_rescaled, predictions_rescaled)
        rmse = sqrt(mse)
        mae = mean_absolute_error(y_test_rescaled, predictions_rescaled)
        r2 = r2_score(y_test_rescaled, predictions_rescaled)

        results[model_name] = [mse, rmse, mae, r2]
    return results
```

This procedure ensures a consistent evaluation of each model across the same dataset, allowing for a fair comparison of performance metrics.

3. RESULTS AND ANALYSIS

3.1 Model Performance Based on RMSE and MAE

The models were evaluated using key error metrics Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) to assess their predictive performance. RMSE measures the square root of the average squared differences between predicted and actual values. As RMSE penalizes larger errors more heavily, it is particularly useful when we want to emphasize the impact of significant deviations in the prediction. On the other hand, MAE gives the average magnitude of the errors without considering their direction, making it easier to interpret but less sensitive to large errors compared to RMSE. The performance of the models across these metrics is visualized in Figure 1, showing that lower error values indicate better model performance. From the analysis of RMSE and MAE, several key observations can be made.

The Linear Regression model produced the highest RMSE value of 0.4496 and an MAE of 0.2950, making it the least accurate model for predicting household energy consumption. Linear regression assumes a linear relationship between the predictors and the target variable, and this assumption likely limits the model's ability to capture the complex, non-linear interactions present in the time series data. Given the dynamic nature

of energy consumption, the poor performance of linear regression was expected. The large RMSE indicates that the model struggles to handle significant deviations between predicted and actual values, while the relatively high MAE confirms that, on average, the model is consistently inaccurate. In contrast, Ridge Regression and Lasso Regression, both of which add regularization to the linear model, showed improved performance over the basic linear regression. Ridge Regression reduced the RMSE to 0.2767 and the MAE to 0.1921, while Lasso Regression achieved an RMSE of 0.2515 and an MAE of 0.2114. The regularization terms in Ridge and Lasso Regression penalize large coefficient values, which mitigates overfitting and improves the model's generalizability. However, while these models perform better than simple linear regression, they are still unable to capture the non-linearity inherent in the dataset. This limitation is reflected in the higher error values compared to more complex models, such as Random Forest and Gradient Boosting.

The Random Forest model achieved the best performance among all models, with an RMSE of 0.2450 and an MAE of 0.1682. Random Forest's strength lies in its ability to model complex interactions by averaging the predictions of multiple decision trees, each trained on different subsets of the data. This aggregation leads to reduced variance and improved generalization, allowing the model to capture non-linear relationships more effectively. Random Forest's low RMSE suggests that it makes fewer large errors, while the low MAE indicates that its predictions are close to the actual values on average. The combination of low RMSE and MAE makes Random Forest the most reliable model in this context.

Table 1. Comparison of Regression Results

Model	MSE	RMSE	MAE	R^2
Linear Regression	0.202169	0.449632	0.295018	-2.920555
Ridge Regression	0.07654	0.276659	0.192069	-0.484305
Lasso Regression	0.063264	0.251524	0.21141	-0.226857
Random Forest	0.060048	0.245046	0.168168	-0.164472
Gradient Boosting	0.084073	0.289954	0.188467	-0.630389
XGBoost	0.069015	0.262707	0.180477	-0.338377
CatBoost	0.058635	0.242147	0.183021	-0.137083
LightGBM	0.08661	0.294296	0.226956	-0.679591

The performance of Gradient Boosting was relatively close to that of Random Forest, with an RMSE of 0.2900 and an MAE of 0.1885. Gradient Boosting builds models sequentially, where each new model corrects the errors made by its predecessors. While Gradient Boosting is generally more flexible than Random Forest, its performance in this case was slightly lower, possibly due to the number of boosting iterations or the choice of hyperparameters. Given more iterations and further tuning, Gradient Boosting has the potential to match or even surpass the performance of Random Forest. However, in its current state, it still shows competitive performance, indicating its strength in handling complex, non-linear data. For the gradient boosting variants, XGBoost and CatBoost demonstrated superior performance compared to the base Gradient Boosting model. XGBoost, with an RMSE of 0.2627 and an MAE of 0.1805, offered a good balance between prediction accuracy and computational efficiency. CatBoost, however, delivered even better results, with an RMSE of 0.2421 and an MAE of 0.1830. CatBoost's advanced handling of categorical features and its optimized gradient boosting algorithm contributed to its superior performance in this time series prediction task. The error values for both XGBoost and CatBoost suggest that these models are highly effective at modeling non-linear interactions, especially in datasets with complex relationships between features.

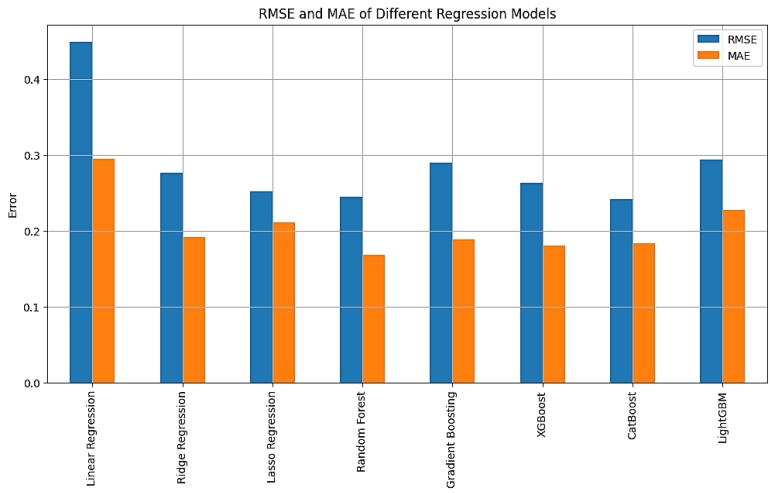

Figure 2. Regression Models Result

Lastly, LightGBM exhibited an RMSE of 0.2943 and an MAE of 0.2270. Although it is known for its speed and efficiency, LightGBM performed slightly worse than CatBoost and XGBoost in this study. The higher RMSE and MAE values suggest that LightGBM might benefit from further tuning to optimize its predictive capabilities. Despite this, LightGBM remains a strong candidate for large-scale datasets where speed is a critical factor, though its accuracy in this particular application falls behind the other gradient boosting methods. The detailed error metrics for all models are summarized in Table 1, which reinforces the observation that ensemble models, especially Random Forest and gradient boosting variants, consistently outperform linear models.

4.2 R^2 Score Analysis

The R^2 score evaluates the proportion of variance in the target variable that is explained by the model. A higher R^2 value (closer to 1) indicates better explanatory power. However, all models in this study produced negative R^2 values, indicating that they performed worse than a simple mean predictor for the data. The Linear Regression model had the worst R^2 score, with a value of -2.9206. This result demonstrates the model's inability to capture the underlying patterns in the data, leading to highly inaccurate predictions. The negative R^2 implies that using the mean of the target variable as a predictor would have been more effective than the model itself. Both Ridge Regression and Lasso Regression produced moderate improvements in their R^2 scores, with values of -0.4843 and -0.2269, respectively. While these regularized models reduced the prediction error compared to the basic linear regression, their ability to explain the variance in the data remains limited. This is likely because these models still assume a linear relationship between the features and the target, which is not well-suited to this complex, non-linear dataset.

Among the ensemble models, Random Forest and CatBoost had the best R^2 scores, at -0.1645 and -0.1371, respectively. Although still negative, these scores suggest that these models were able to capture some of the variance in the data, though not enough to outperform a simple baseline predictor. The relatively higher R^2 values for these models compared to the linear models indicate their ability to capture more of the underlying complexity in the data. The negative R^2 values across all models highlight the challenging nature of the dataset. Energy consumption is influenced by many external factors, some of which may not be fully captured in the current feature set. Furthermore, the noise present in the data may also be contributing to the models' inability to explain the variance effectively. Achieving positive R^2 values would likely require further feature engineering, incorporating additional external variables (such as weather patterns or socio-economic factors), and more sophisticated model architectures.

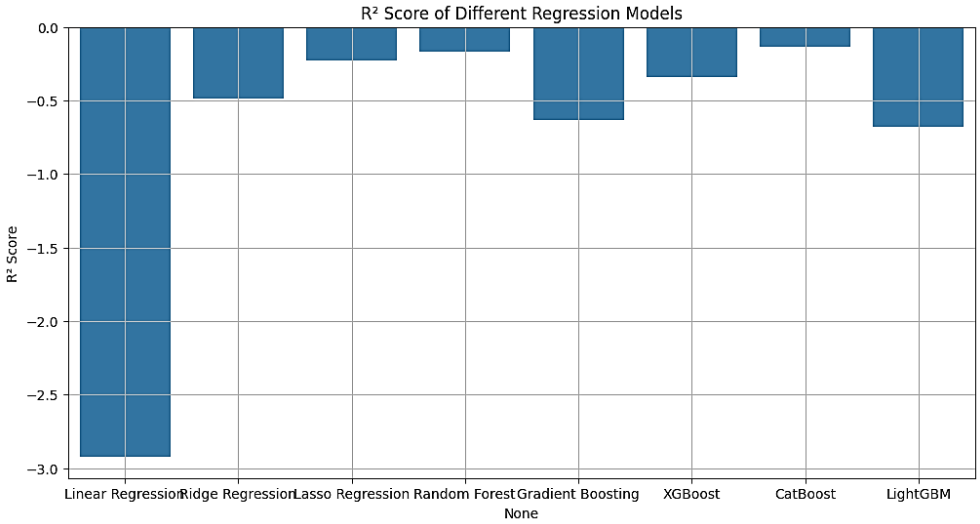

Figure 3. R2 Score of Regression Model Result

4.3. Discussion and Recommendations

The analysis of the results reveals several important insights regarding the suitability of different models for time series prediction of household energy consumption. Firstly, the ensemble methods, especially Random Forest and CatBoost, consistently outperformed the linear models. This reinforces the importance of capturing non-linear relationships in time series data, where simple linear models fall short. Random Forest, with its ability to model complex interactions between features, demonstrated the best performance overall, making it a strong choice for energy consumption prediction tasks. CatBoost, with its specialized gradient

boosting algorithm, also performed well and should be considered when computational efficiency is critical. Secondly, while regularized models like Ridge Regression and Lasso Regression showed improvements over basic linear regression, their inability to model non-linearity limited their effectiveness. Regularization is useful in preventing overfitting, but it cannot compensate for the lack of flexibility in linear models when dealing with complex datasets.

The results also suggest that gradient boosting variants such as XGBoost and CatBoost are highly competitive models for this task. Their superior performance compared to basic Gradient Boosting highlights the advantages of using more advanced boosting algorithms that incorporate techniques like regularization and efficient handling of categorical variables. Finally, the negative R^2 values for all models indicate that there is room for improvement in the modeling approach. Further feature engineering, including the integration of external factors like weather conditions or economic indicators, could potentially enhance model performance. Additionally, more complex models, such as recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks, should be explored as they are well-suited to time series data and could potentially yield better results.

4.4 Practical Applications

The findings of this study offer several practical applications for implementing predictive models in real-world household energy management systems. For instance, the Random Forest model, which demonstrated the best overall performance, can be integrated into smart home energy management systems to optimize energy usage dynamically. By predicting daily energy consumption, homeowners can receive actionable insights into their energy habits, enabling them to adjust appliance usage during peak hours to reduce costs. Additionally, these models can serve as a foundation for demand-side management programs in smart grids. Utility companies can leverage accurate energy predictions to balance supply and demand more efficiently, reducing the risk of power outages and enhancing grid stability. Furthermore, integrating the model predictions with renewable energy sources such as solar panels can enhance the reliability and sustainability of residential energy systems by aligning energy storage and distribution with anticipated consumption patterns.

From a commercial perspective, these predictive models can be employed by energy service providers to create personalized energy plans for consumers. These plans could be tailored based on historical usage and predicted future demand, encouraging energy-saving behaviors and promoting the adoption of renewable energy solutions. In rural areas with limited access to consistent electricity, predictive analytics could optimize microgrid operations, ensuring that energy resources are allocated effectively based on forecasted demand. Finally, these models hold potential for deployment in smart city initiatives, where energy consumption data from various households can be aggregated to inform city-wide energy policies. The integration of such machine learning-driven predictions can pave the way for more sustainable urban planning, focusing on minimizing energy wastage and promoting green energy adoption.

4. CONCLUSION

This study evaluated the performance of several machine learning regression models for predicting household energy consumption. The models tested include traditional linear regression methods, regularized versions such as Ridge and Lasso regression, and more advanced ensemble methods like Random Forest, Gradient Boosting, XGBoost, CatBoost, and LightGBM. Each model was evaluated based on key metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2. The results show a clear distinction in performance between the linear models and the ensemble-based methods. Linear Regression performed the worst, with the highest RMSE and MAE values, demonstrating its limitations in capturing the complex, non-linear patterns inherent in time series data. Even with regularization, both Ridge Regression and Lasso Regression could not outperform the more sophisticated ensemble models.

Among the ensemble models, Random Forest emerged as the best performer, achieving the lowest MAE and competitive RMSE. This model's ability to average the predictions of multiple decision trees resulted in reduced prediction errors and improved generalization, making it a reliable model for energy consumption forecasting. Gradient Boosting also delivered strong performance, though slightly behind Random Forest in terms of error metrics. When compared with the standard Gradient Boosting model, advanced versions like XGBoost and CatBoost exhibited further improvements in prediction accuracy, with CatBoost marginally outperforming XGBoost. The R^2 scores, while negative across all models, reflect the difficulty of the task. Negative R^2 values indicate that the models performed worse than a simple baseline prediction using the mean of the target variable. This result underscores the complexity of the dataset, which likely includes external factors and noise that are difficult to model accurately. Nevertheless, ensemble methods such as Random Forest and CatBoost still demonstrated an ability to capture a portion of the variance, making them the best candidates for future improvements.

In conclusion, non-linear models, especially those based on ensemble methods, are more suitable for predicting household energy consumption compared to linear regression models. Random Forest and CatBoost were identified as the top-performing models in this study, but further improvements could be achieved by exploring more sophisticated techniques like deep learning models, including recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks. Additionally, incorporating external factors, such as weather data or socio-economic variables, could enhance the model's ability to explain more of the variance and improve overall prediction accuracy. Future work should focus on further tuning of ensemble models, considering deep learning architectures for handling time series data, and expanding the feature set to include more external influences on energy consumption. Such steps could potentially overcome the challenges observed in this study and yield more accurate and robust predictions.

REFERENCES

[1] S. K. Rathor and D. Saxena, "Energy management system for smart grid: An overview and key issues," *Int. J. Energy Res.*, vol. 44, no. 6, pp. 4067–4109, 2020.

[2] M. S. Aliero, K. N. Qureshi, M. F. Pasha, and G. Jeon, "Smart Home Energy Management Systems in Internet of Things networks for green cities demands and services," *Environ. Technol. \& Innov.*, vol. 22, p. 101443, 2021.

[3] J. Leitao, P. Gil, B. Ribeiro, and A. Cardoso, "A survey on home energy management," *IEEE Access*, vol. 8, pp. 5699–5722, 2020.

[4] A. Raza, L. Jingzhao, Y. Ghadi, M. Adnan, and M. Ali, "Smart home energy management systems: Research challenges and survey," *Alexandria Eng. J.*, vol. 92, pp. 117–170, 2024.

[5] U. Mir, U. Abbasi, T. Mir, S. Kanwal, and S. Alamri, "Energy management in smart buildings and homes: current approaches, a hypothetical solution, and open issues and challenges," *IEEE Access*, vol. 9, pp. 94132–94148, 2021.

[6] S. Mohseni, A. C. Brent, S. Kelly, and W. N. Browne, "Demand response-integrated investment and operational planning of renewable and sustainable energy systems considering forecast uncertainties: A systematic review," *Renew. Sustain. Energy Rev.*, vol. 158, p. 112095, 2022.

[7] C. Ntakolia, A. Anagnostis, S. Moustakidis, and N. Karcanias, "Machine learning applied on the district heating and cooling sector: A review," *Energy Syst.*, vol. 13, no. 1, pp. 1–30, 2022.

[8] A. A. A. Gassar and S. H. Cha, "Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review," *Energy Build.*, vol. 224, p. 110238, 2020.

[9] M. M. Hussain, R. Akram, Z. A. Memon, M. H. Nazir, W. Javed, and M. Siddique, "Demand side management techniques for home energy management systems for smart cities," *Sustainability*, vol. 13, no. 21, p. 11740, 2021.

[10] N. C. Iraganaboina and N. Eluru, "An examination of factors affecting residential energy consumption using a multiple discrete continuous approach," *Energy Build.*, vol. 240, p. 110934, 2021.

[11] S. F. Ahmed *et al.*, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13521–13617, 2023.

[12] T. Ahmad, H. Zhang, and B. Yan, "A review on renewable energy and electricity requirement forecasting models for smart grid and buildings," *Sustain. Cities Soc.*, vol. 55, p. 102052, 2020.

[13] N. Mlilo, J. Brown, and T. Ahfock, "Impact of intermittent renewable energy generation penetration on the power system networks--A review," *Technol. Econ. Smart Grids Sustain. Energy*, vol. 6, no. 1, p. 25, 2021.

[14] S. Ben Atitallah, M. Driss, W. Boulila, and H. Ben Ghézala, "Leveraging Deep Learning and IoT big data analytics to support the smart cities development: Review and future directions," *Comput. Sci. Rev.*, vol. 38, p. 100303, 2020.

[15] T. Alam and A. AlArjani, "A Comparative Study of CO2 Emission Forecasting in the Gulf Countries Using Autoregressive Integrated Moving Average, Artificial Neural Network, and Holt-Winters Exponential Smoothing Models," *Adv. Meteorol.*, vol. 2021, no. 1, p. 8322590, 2021.

[16] E. T. Winn-Nuñez, M. Griffin, and L. Crawford, "A simple approach for local and global variable importance in nonlinear regression models," *Comput. Stat. \& Data Anal.*, vol. 194, p. 107914, 2024.

[17] I. K. Ofori, C. K. Obeng, and S. A. Asongu, "What really drives economic growth in sub-Saharan Africa? Evidence from the lasso regularization and inferential techniques," *J. Knowl. Econ.*, vol. 15, no. 1, pp. 144–179, 2024.

[18] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," *Geosci. Front.*, vol. 12, no. 1, pp. 469–477, 2021.

[19] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian, "A critical review of wind power forecasting methods—past, present and future," *Energies*, vol. 13, no. 15, p. 3764, 2020.

[20] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interact. Learn. Environ.*, vol. 31, no. 6, pp. 3360–3379, 2023.

[21] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J. big data*, vol. 7, no. 1, p. 94, 2020.

[22] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, "The future of forecasting for renewable energy," *Wiley Interdiscip. Rev. Energy Environ.*, vol. 9, no. 2, p. e365, 2020.

[23] J. Zheng, J. Du, B. Wang, J. J. Klemeš, Q. Liao, and Y. Liang, "A hybrid framework for forecasting power generation of multiple renewable energy sources," *Renew. Sustain. Energy Rev.*, vol. 172, p. 113046, 2023.

[24] O. Erixno, N. Abd Rahim, F. Ramadhani, and N. N. Adzman, "Energy management of renewable energy-based combined heat and power systems: A review," *Sustain. Energy Technol. Assessments*, vol. 51, p. 101944, 2022.

[25] M. Khalil, A. S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption—A systematic review," *Eng. Appl. Artif. Intell.*,

vol. 115, p. 105287, 2022.
[26] H. Yesilyurt, Y. Dokuz, and A. S. Dokuz, "Data-Driven Energy Consumption Prediction of a University Office Building using Machine Learning Algorithms," *Energy*, p. 133242, 2024.
[27] L. Zhang and D. Jánoš\'\ik, "Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches," *Expert Syst. Appl.*, vol. 241, p. 122686, 2024.
[28] Y. Liang, S. Li, C. Yan, M. Li, and C. Jiang, "Explaining the black-box model: A survey of local interpretation methods for deep neural networks," *Neurocomputing*, vol. 419, pp. 168–182, 2021.
[29] T. V. Singh, "Smart Home Dataset with Weather Information." 2021.

BIBLIOGRAPHY OF AUTHORS

Gregorius Airlangga, Received the B.S. degree in information system from the Yos Sudarso Higher School of Computer Science, Purwokerto, Indonesia, in 2014, and the M.Eng. degree in informatics from Atma Jaya Yogyakarta University, Yogyakarta, Indonesia, in 2016. He got Ph.D. degree with the Department of Electrical Engineering, National Chung Cheng University, Taiwan. He is also an Assistant Professor with the Department of Information System, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia. His research interests include artificial intelligence and software engineering include path planning, machine learning, natural language processing, deep learning, software requirements, software design pattern and software architecture.