Decoding Energy Usage Predictions: An Application of XAI Techniques for Enhanced Model Interpretability

[1]**Gregorius Airlangga**
[1]Information System Study Program, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: [1]gregorius.airlangga@atmajaya.ac.id

Article Info

Article history:
Received Jan 4th, 2024
Revised Mar 26th, 2024
Accepted Apr 26th, 2024

Keyword:
Energy Usage Predictions
SHAP LIME
XAI
XGBoost

ABSTRACT

The growing complexity of machine learning models has heightened the need for interpretability, particularly in applications impacting resource management and sustainability. This study addresses the challenge of interpreting predictions from sophisticated machine learning models used for building energy consumption predictions. By leveraging Explainable AI (XAI) techniques, including Permutation Importance, SHapley Additive exPlanations (SHAP), and Local Interpretable Model-Agnostic Explanations (LIME), we have dissected the predictive features influencing building energy usage. Our research delves into a dataset consisting of various building characteristics and weather conditions, applying an XGBoost model to predict Site Energy Usage Intensity (Site EUI). The Permutation Importance method elucidated the global significance of features across the dataset, while SHAP provided a dual perspective, revealing both the global importance and local impact of features on individual predictions. Complementing these, LIME offered rapid, locally focused interpretations, showcasing its utility for instances where immediate insights are essential. The findings indicate that 'Energy Star Rating', 'Facility Type', and 'Floor Area' are among the top predictors of energy consumption, with environmental factors also contributing to the models' decisions. The application of XAI techniques yielded a nuanced understanding of the model's behavior, enhancing transparency and fostering trust in the predictions. This study contributes to the field of sustainable energy management by demonstrating the application of XAI for insightful model interpretation, reinforcing the significance of interpretable AI in the development of energy policies and efficiency strategies. Our approach exemplifies the balance between predictive accuracy and the necessity for model transparency, advocating for the continued integration of XAI in AI-driven decision-making processes.

Corresponding Author:
Gregorius Airlangga,
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.v7i2.29041

1. INTRODUCTION

In the contemporary landscape of data-driven decision-making, the task of accurately predicting building energy usage stands at the forefront of efforts to enhance energy management, drive sustainability, and inform policy making [1]–[3]. The Women in Data Science (WiDS) Datathon 2022 encapsulates this challenge, presenting an intriguing problem: predicting the Site Energy Usage Intensity (Site EUI) of buildings across various states in the United States [4]–[6]. This problem is addressed using a comprehensive dataset encompassing roughly 100,000 observations collected over seven years, incorporating an array of variables that range from specific building characteristics, like floor area and facility type, to location-centric weather data, such as annual average temperature and total precipitation [7]–[9]. The significance of this task goes beyond the mere statistical prediction of energy usage. It delves into the realm of Explainable AI (XAI), a field rapidly gaining traction in the AI community [10]–[12]. XAI stands as a cornerstone in this context, offering not just transparency but also crucial insights into the decision-making processes of machine learning models. In an era where global climate challenges are intensifying and the call for sustainable energy solutions is becoming increasingly urgent, the ability to predict building energy usage with precision and understanding is more vital than ever. This urgency is further amplified by the rapid pace of urbanization, leading to an escalation in building-related energy demands [13]–[15].

The core challenge in deploying machine learning models for such predictions lies in their complexity and the accompanying lack of transparency. Models based on deep learning, for instance, though capable of achieving high levels of precision, often operate as 'black boxes' [16]–[18]. Their inherent complexity renders them opaque, making it difficult for stakeholders to understand and trust their decision-making processes. This lack of transparency is not just a technical issue but also a matter of ethical concern, especially in sectors with significant societal impacts like energy management. It is here that XAI becomes invaluable, serving as a bridge between the high accuracy of complex models and the necessity for their decisions to be understandable and ethically sound. The existing work in the field of building energy consumption prediction is both extensive and diverse, encompassing a multitude of modeling approaches. Early studies primarily focused on linear regression and time-series forecasting techniques. While these methods offer a degree of interpretability, they often struggle to effectively manage complex, non-linear relationships inherent in large-scale datasets. The advent of deep learning brought about models with enhanced predictive accuracy but at the cost of reduced interpretability. This trade-off between model complexity and interpretability is a recurring theme in predictive modeling research and forms the crux of the current study.

XAI emerges as a potent solution to this predicament, aiming to demystify the decision-making processes of AI models. The importance of XAI is particularly underscored in studies that highlight its potential in enhancing trust and facilitating more informed decision-making across various domains [19]–[21]. Despite its critical importance, the application of XAI techniques in the domain of energy prediction remains largely underexplored. This research gap is particularly glaring given the global emphasis on sustainable energy practices and the need for transparent decision-making processes that are accessible and understandable, even to non-expert stakeholders. Addressing this research gap, our study makes a novel contribution by applying three advanced XAI techniques: Permutation Importance [22], SHAP (SHapley Additive exPlanations) [23], and LIME (Local Interpretable Model-Agnostic Explanations) [24], to the domain of building energy usage prediction. Permutation Importance, implemented using the ELI5 library, offers a straightforward approach to assess the significance of different features by observing the impact of their random permutation on model accuracy. SHAP provides a more nuanced understanding by decomposing model predictions into individual feature contributions. LIME, on the other hand, enhances the interpretability of predictions at a local, individual observation level.

Our methodology encompasses a comprehensive analysis of the dataset, including meticulous handling of missing values, nuanced feature engineering, and data normalization, all culminating in the application of an eXtreme Gradient Boosting (XGBoost) regression model. The choice of XGBoost is motivated by its proven effectiveness in managing large datasets and its compatibility with the XAI methods employed. This approach not only aims to enhance the predictive accuracy of the models but also provides a detailed understanding of how different building characteristics and environmental factors collectively influence energy consumption predictions. The subsequent sections of the paper are meticulously crafted to build upon this foundation. The methodology section details the data preprocessing steps, feature engineering, the development of the XGBoost predictive model, and the rationale behind the selection of the XAI techniques. The results section presents the findings from the application of the XGBoost model, including its predictive performance and the insights gleaned from the XAI methods, especially regarding the most influential factors in predicting building energy usage. The discussion section interprets these results within the broader context of existing literature, examining the implications of our findings for energy management and policymaking, and acknowledging the study's limitations while suggesting avenues for future research. Finally, the conclusion summarizes the key contributions of the study, emphasizing the critical role of XAI in building energy prediction and its potential impact on the landscape of sustainable energy management.

2. RELATED THEORY

The foundation of this research lies in the intersection of predictive modeling, energy efficiency, and XAI, each contributing critical theoretical underpinnings to our study. This section delves into the related theories that inform our research, providing a backdrop against which our methodology and findings can be contextualized.

2.1. Predictive Modeling in Energy Efficiency

Predictive modeling in the realm of energy efficiency revolves around the use of statistical and machine learning techniques to forecast energy consumption patterns. This aspect of the research is grounded in the theory of regression analysis, a statistical method used for predicting a continuous dependent variable based on one or more independent variables. The application of regression in energy prediction has evolved from simple linear models to more complex ones like the XGBoost algorithm used in this study. XGBoost [25], an advanced form of gradient boosting, is based on the principle of ensemble learning, where multiple models (often called "weak learners") are trained and combined to improve the robustness and accuracy of predictions. Predictive modeling in energy efficiency relies heavily on regression analysis, which can be represented mathematically. In its simplest form, linear regression is modeled as presented in the equation (1).

$$\mathrm{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{1}$$

where Y is the dependent variable (energy consumption), X_i are independent variables (building characteristics, weather conditions), β_i are coefficients, and ϵ is the error term. The XGBoost algorithm, a more advanced model used in our study, operates on the principle of gradient boosting, which involves the sequential addition of predictors that correct the predecessors' errors. Mathematically, the prediction y_i at the *i*-th iteration is presented in the equation (2).

$$\widehat{y_i^{(i)}} = \widehat{y_i^{(i-1)}} + \eta \cdot h_i(X) \tag{2}$$

Where $h_i(X)$ is the output of the new model added at the *i*-th iteration, and η is the learning rate.

2.2. Explainable AI (XAI)

The concept of Explainable AI emerges from the need to make AI systems more transparent and understandable to humans [11]. XAI is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. In our study, XAI is crucial for providing insights into the predictive models used for estimating building energy consumption. Theories underpinning XAI emphasize the importance of interpretability and transparency, especially in models that are otherwise considered 'black boxes'. The XAI methods employed in this research – Permutation Importance, SHAP, and LIME – each have their theoretical foundations. Permutation Importance is based on the idea of feature importance ranking through random permutations, SHAP values are derived from game theory and offer a consistent approach to feature attribution, it can be calculated by equation (3).

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f_x(S \cup \{i\}) - f_x(S)] \tag{3}$$

Where *N* is the set of all features, *S* is a subset of features excluding *i* and $f_x(S)$ is the prediction for the subset *S*. The SHAP value ϕ_i represents the average marginal contribution of a feature *i* across all possible combinations. In addition, LIME's theory revolves around creating interpretable models that approximate the predictions of complex models locally. It approximates the local behavior of complex models using simpler models such as linear model as presented in equation (4).

$$g(z') = \beta_0 + \sum \beta_i z_i' \tag{4}$$

Where $g(z')$ is the explanation model, z_i' are the interpretable representations of the features, and b_i are the coefficients learned by the explanation model.

2.3. Energy Efficiency and Building Characteristics

The theory linking building characteristics to energy efficiency is a critical aspect of this research. This theory posits that various attributes of a building, such as its size, age, design, and construction materials, significantly impact its energy consumption patterns [26]. Weather factors, including temperature and humidity, also play a crucial role in determining a building's energy usage. Understanding these relationships is essential for developing accurate predictive models in the field of energy efficiency. The relationship between building characteristics and energy efficiency is often explored through regression models, where energy consumption (EC) is a function of various building attributes (BA) as presented in the equation (5).

$$EC = f(BA_1, BA_2, \dots, BA_n) \tag{5}$$

2.4. Statistical Learning Theory

Statistical learning theory provides the mathematical framework for machine learning. It involves understanding how different algorithms learn from data to make predictions or decisions [27]. This theory is particularly relevant to our research as it guides the model selection, validation, and testing processes. Concepts such as bias-variance trade-off, overfitting, and underfitting are integral to understanding the performance and limitations of the predictive models used in this study. A core concept in statistical learning is the bias-variance trade-off, which is crucial in model selection. The expected prediction error of a model can be decomposed as presented in the equation 6.

$$Error = Bias^2 + Variance + Irreducible\ Error \quad (6)$$

3. RESEARCH METHOD

3.1. Dataset and Data Preprocessing

This study utilizes a meticulously collated dataset from the WiDS Datathon 2022, featuring approximately 100,000 observations of building energy usage across various states in the United States, collected over seven years [28]. This dataset is rich in variables, ranging from specific building characteristics like floor area, building class, and facility type, to granular weather data, including metrics such as annual average temperature and total precipitation. The initial stage of data preprocessing involved a thorough cleaning process. Data integrity was paramount; hence, meticulous steps were taken to identify and rectify any inconsistencies or inaccuracies. This included addressing outliers – data points that deviate significantly from the majority of the data – which were either corrected or removed to prevent potential skewing of the predictive models. Missing values posed a significant challenge, given their potential to impact the model's performance adversely. A strategy was employed to handle missing data based on the nature of the variables. For continuous variables, missing values were imputed with the median or mean, as these measures provide a central tendency of the data that is less sensitive to outliers. For categorical variables, we used the mode for imputation, replacing missing values with the most frequently occurring category in the dataset. Normalization was another critical step in our preprocessing phase. Given the diversity of the variables in terms of scale and units (e.g., square footage of floor area vs. annual precipitation in millimeters), normalization was essential to ensure that no single feature disproportionately influenced the model's predictions. This process involved scaling all numerical features to a standard range, which not only facilitated a fair comparison across different variables but also expedited the convergence of the machine learning algorithm during training.

3.2. Feature Engineering

Feature engineering is a crucial process of transforming raw data into meaningful features that significantly improve the efficacy of machine learning models. This process was driven by both domain expertise and exploratory data analysis. For instance, we derived a 'building age' feature from the 'year built' data, providing a more direct measure of a building's age, which could be more relevant for energy usage predictions than the construction year alone. We also synthesized climate indicators by aggregating various weather parameters, such as combining temperature and humidity levels to create a more comprehensive 'comfort index'.

3.3. Model Development: XGBoost Regression Model

The XGBoost regression model was chosen for its robustness and effectiveness in handling diverse and large datasets. XGBoost, an implementation of gradient-boosted decision trees, is designed for speed and performance and is particularly well-suited for the high-dimensional and feature-rich dataset in our study. The model development involved partitioning the dataset into a training set and a test set, with the training set used to train the model and the test set reserved for evaluating its predictive performance. A critical aspect of this phase was hyperparameter tuning, where we experimented with various combinations of parameters like learning rate, depth of trees, and the number of trees to find the optimal configuration for our model. This fine-tuning was instrumental in enhancing the model's ability to generalize well to new, unseen data, striking a delicate balance between underfitting and overfitting.

3.4. Application of Explainable AI (XAI) Techniques

Integrating XAI techniques into our study was a strategic decision aimed at enhancing the interpretability of our model's predictions. We began with the application of permutation importance, utilizing the ELI5 library. This approach involved shuffling individual features within the dataset and observing the resultant impact on the model's performance. By evaluating the changes in performance metrics, we were able to quantify the significance of each feature, providing a concrete measure of its influence on the accuracy of the model. Following this, we employed SHAP values, rooted in cooperative game theory, to dissect and elucidate the contribution of each feature to individual predictions. This process entailed breaking down each prediction into the cumulative effects of each feature's introduction to the model, offering a thorough and nuanced perspective on how each feature affects the model's output. Additionally, we incorporated LIME to interpret individual predictions. LIME accomplishes this by constructing simple, local surrogate models around each prediction, shedding light on how various features in a specific instance contribute to the final prediction. This technique allows for a detailed understanding of the model's behavior at the level of individual predictions, providing valuable insights into the inner workings of our predictive model.

3.5. Model Evaluation and Validation

Evaluating and validating the performance of our model was a multifaceted process. We used standard evaluation metrics such as Mean Squared Error (MSE) to quantify the model's accuracy and R-squared to determine the proportion of variance in the dependent variable that was predictable from the independent variables. To ensure the model's robustness and generalizability, we employed k-fold cross-validation, a technique where the dataset is divided into k subsets, and the model is trained and tested k times, each time with a different subset as the test set and the remaining data as the training set. This method provided a comprehensive assessment of the model's performance across different subsets of data, reducing the likelihood of anomalies influencing the results.

3.6. Ethical Considerations and Data Privacy

Throughout this research, ethical considerations and data privacy were paramount. The dataset was anonymized to protect privacy and confidentiality, ensuring that no sensitive or personally identifiable information was accessible. The study was conducted with strict adherence to ethical research standards and data protection regulations, reflecting our commitment to responsible and ethical data handling and analysis.

4. RESULTS AND ANALYSIS

In contemporary research, where machine learning (ML) models play a pivotal role in decision-making processes, the quest for precision often overshadows the need for transparency and understandability. This is particularly evident in the realm of deep learning models, which, while achieving remarkable levels of accuracy, frequently operate as enigmatic 'black boxes'. The imperative, therefore, is not just for these models to perform efficiently but also for their decision-making mechanisms to be transparent and comprehensible. It is essential that we, as humans, can fully grasp how decisions are made by AI systems to foster trust and reliance on these technologies. Transparent and explainable ML models are not just a preference but a necessity for their practical application and acceptance. In this research, we address this crucial need for explainability in ML through a detailed exploration of three significant model explainability methods: Permutation Importance, SHAP, and LIME. Each of these methods offers unique insights into the internal workings of ML models, providing us with the tools to delve into the 'why' and 'how' behind the predictions made by our models.

Our primary task involves predicting the energy consumption of buildings, a task for which our model has been trained and tested. While the model adeptly predicts energy consumption for each sample in the test data, the core of our investigation revolves around deciphering the rationale behind these predictions. We seek to answer critical questions such as: Why is our model predicting the values it predicts? Which variables are positively or negatively correlated with the target? And importantly, which variables hold the most significant sway in the prediction process, either for the dataset as a whole or for individual examples? The integration of Permutation Importance, SHAP, and LIME into our analytical framework allows us to methodically address these queries. Permutation Importance helps us identify the most influential features by observing changes in the model's accuracy when the values of each feature are randomly shuffled.

SHAP, rooted in cooperative game theory, provides a more granular understanding by attributing each prediction to the contribution of individual features. LIME complements these by offering local interpretations, explaining predictions on a per-instance basis, and thus illuminating how specific combinations of features drive the model's output. This comprehensive approach not only enhances our understanding of the model's predictive behavior but also demystifies the complex interactions between various features and their impact on the final predictions. By employing these methods, our study transcends beyond mere predictive accuracy, venturing into the realms of accountability and interpretability in machine learning. This endeavor aligns closely with the growing demand for ethical AI, where the ability to explain and justify algorithmic decisions is as crucial as the decision-making process itself.

4.1. Permutation Importance Result

Based on table 1 that describing about the Permutation Importance results, the analysis indicates that the 'Energy Star Rating' is the most influential feature when predicting the energy consumption of a building within the dataset used. This suggests that the energy efficiency rating assigned to a building, which often encapsulates various aspects of its design and operation, is a significant predictor of its energy usage. Following the 'Energy Star Rating', the 'Facility Type' emerges as the second most important feature. This aligns with the understanding that the kind of activity a building is used for—whether it's a hospital, office, school, or otherwise—has a substantial impact on its energy consumption patterns. Different facilities will inherently have varying energy needs based on their usage patterns, equipment, and occupancy schedules.

Table 1. The Performance of Permutation Importance

Weight	Feature
0.2749 ± 0.0058	facility_star_rating
0.1905 ± 0.0058	energy_type
0.0938 ± 0.0067	floor_area
0.0312 ± 0.0055	building_class
0.0218 ± 0.0061	Scale_Factor
0.0132 ± 0.0073	year_built
0.0095 ± 0.0055	building_avg_temp
0.0091 ± 0.0065	ELEVATION
0.0091 ± 0.0075	january_avg_temp
0.0089 ± 0.0065	february_avg_temp
0.0089 ± 0.0054	days_below_20F
0.0084 ± 0.0054	heating_degree_days

The ‘Floor Area’ of a building comes in as the third most influential factor according to the Permutation Importance results. Intuitively, this makes sense since larger buildings typically have greater energy needs for heating, cooling, and lighting compared to smaller ones. The weights assigned to these features, accompanied by their standard deviations, provide a quantified measure of their importance. The weight represents the feature’s average impact on the accuracy of the model when its values are randomly shuffled, thus breaking the link between the feature and the outcome. The standard deviation gives an indication of the variability of the feature’s weight, which can be seen as a measure of the confidence in the importance ranking. To interpret these results in the context of energy consumption prediction, it can be concluded that initiatives aimed at improving a building’s energy star rating could potentially lead to substantial gains in energy efficiency. Similarly, understanding the energy profiles of different facility types can inform targeted energy-saving strategies. Lastly, the impact of floor area on energy consumption highlights the importance of space management and optimization in energy conservation efforts.

4.2. SHAP Result

The visual provided in figure 1 appears to be a SHAP summary plot, which illustrates the impact of each feature on the model’s output. The SHAP summary plot visualizes the mean absolute SHAP values for each feature across all the samples, which gives us an understanding of the overall importance of each feature in the model. In the given SHAP summary plot, the features are ranked by the mean absolute SHAP value, which is indicated on the x-axis. This value quantifies the average impact of a feature on the model’s output magnitude, disregarding the direction of the impact (i.e., whether it increases or decreases the prediction). The features with larger mean absolute SHAP values are deemed to have a greater influence on the model’s predictions.

The ‘energy_star_rating’ emerges as the most significant feature, with the largest mean absolute SHAP value, indicating that it has the most substantial average effect on the energy consumption predictions. This is followed by ‘facility_type’ and ‘building_class’, which also exhibit high mean absolute SHAP values, showing they are important predictors in the model. These top features suggest that the model gives considerable weight to the energy efficiency rating, the type of facility, and the building class when estimating energy usage. Features like ‘State_Factor’, ‘floor_area’, and ‘year_built’ also contribute notably to the model’s predictions, as indicated by their positive mean SHAP values. These factors are likely related to geographical influences, physical size, and age of the building, respectively, each playing a significant role in determining energy consumption.

The variables 'january_avg_temp', 'january_min_temp', and 'snowdept’_inches' suggest that weather conditions during winter, particularly in January, and snow depth are also influential factors, although to a lesser degree than the top features. The plot also shows the summed impact of 51 other features, which have smaller individual mean absolute SHAP values. This indicates that while these features do contribute to the model’s predictions, their influence is less pronounced when compared to the leading features. The key takeaway from the SHAP summary plot is that while all the listed features influence the model’s output, the magnitude of their impact varies. Understanding these effects allows researchers and practitioners to focus on the most influential factors when considering building energy efficiency interventions and can also provide insights into areas where data collection and feature engineering could be improved. In practice, these results would support targeted strategies to improve energy efficiency. For example, efforts could be focused on improving the ‘energy_star_rating’ of buildings, or specific adjustments could be made for certain ‘facility_types’ or ‘building_classes’ known to have higher energy consumption. Furthermore, the influence of weather-related features underscores the importance of considering seasonal and climate factors in energy usage predictions and planning.

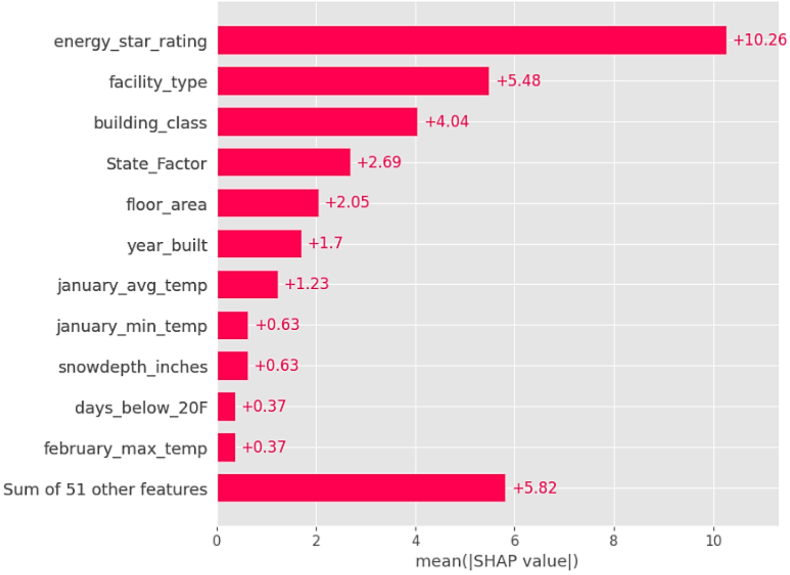


Figure 1. SHAP Result

4.3. LIME Result

As presented in the figure 2, the LIME method has been employed to deconstruct a particular prediction from a complex XGBoost model, offering a granular and intuitive understanding of the model's behavior for an individual instance. This local explanation is crucial for identifying the direct influence of specific features on the model's output, thereby enabling a level of interpretability that is often lacking in sophisticated machine learning models. In the LIME visualization presented, each feature is associated with a value range, and its influence is denoted by the length and direction of a bar—green for a positive contribution and red for a negative impact on the predicted outcome. The visualization indicates that the 'Energy Star Rating' for this instance, when exceeding a threshold of 0.65, appears to have a substantial and positive influence on the energy consumption prediction, which is depicted by a prominent green bar. This might initially seem paradoxical as energy star ratings are generally indicative of energy efficiency; however, in this context, it could be reflective of an underlying pattern where buildings with higher energy star ratings—potentially due to larger sizes or more advanced functionalities—also exhibit higher energy usage.

Conversely, the 'Facility Type' feature, when below a threshold of 0.09, has been shown to contribute negatively to energy consumption, as demonstrated by a significant red bar. This suggests that types of facilities, as categorized within this dataset, are predicted to consume less energy. Additionally, the 'Building Class' feature negatively correlates with the energy prediction, indicating that certain classifications of buildings are associated with lower energy consumption within the model's reasoning for this specific case. Interestingly, the feature 'Floor Area' presents a divided impact with both positive and negative contributions within a specified range, implying a nuanced relationship where the floor area by itself does not unambiguously affect the energy consumption prediction for this instance. This might indicate that the effect of floor area on energy consumption is conditional on other factors or that it has a complex interaction with other variables not immediately discernible from the global model.

The visualization also highlights other features like 'State Factor' and 'Year Built', which exert a smaller influence on the prediction. Weather-related features, such as 'snowdepth_inches' and 'days_above_110F', underscore the role of environmental conditions, pointing towards an increase in energy usage for heating or cooling in response to weather extremes. The core strength of LIME lies in its local fidelity—the linear model it constructs to approximate the predictions of the complex model is particularly accurate around the specific instance under analysis. This is achieved by generating synthetic samples in the vicinity of the instance and using the predictions from the original model, weighted by their proximity, to build an interpretable linear model. The outcome is a locally faithful explanation, as evidenced by the proximity-focused interpretation, which allows for a high degree of confidence in understanding why the model made a specific prediction. Through LIME, we gain insights that are not just relevant but also actionable. For example, in the realm of building energy management, understanding the positive correlation between a high energy star rating and increased energy usage for a particular building can guide energy efficiency initiatives. Similarly, recognizing how facility type influences energy consumption can lead to more targeted and effective energy conservation strategies.

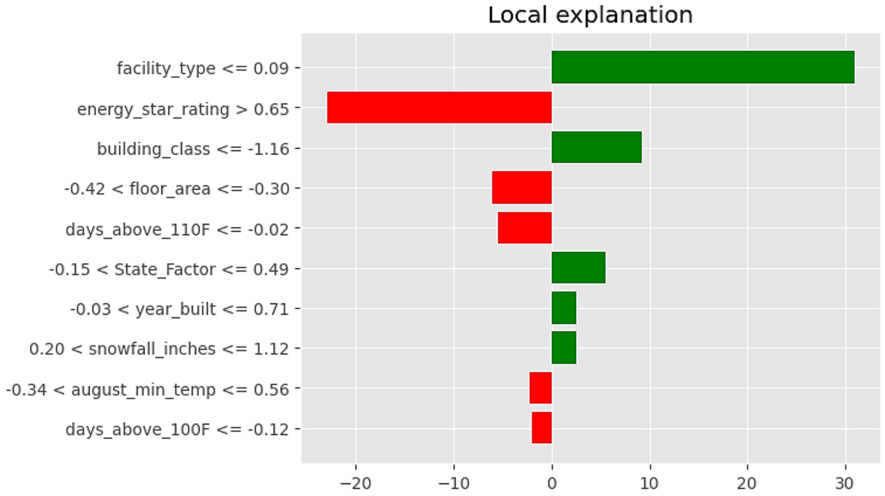
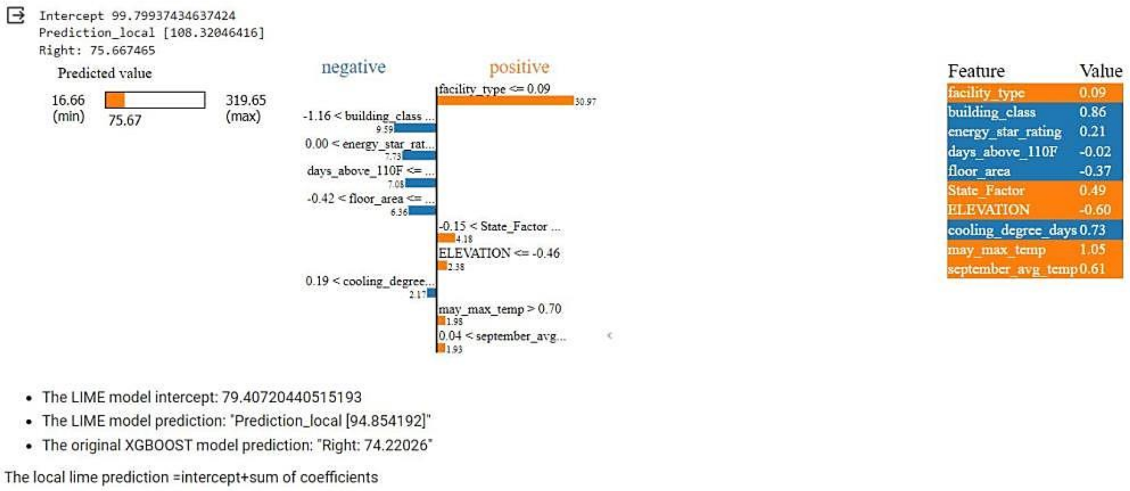


Figure 2. LIME Result

From the figure 3, we can interpret that for the given instance, 'facility_type' has the most significant positive impact on the model's prediction, suggesting that when the facility type is less than or equal to 0.09, it considerably increases the predicted value. Conversely, 'building_class' and 'floor_area' show a strong negative influence, indicating that certain values for these features lead to a lower predicted outcome. Additionally, features like 'State_Factor' and 'ELEVATION' appear to have a smaller negative impact. In contrast, 'cooling_degree_days', 'may_max_temp', and 'september_avg_temp' have a positive influence but with a lesser magnitude compared to 'facility_type'.



```
Intercept 99.79937434637424
Prediction_local [108.32046416]
Right: 75.667465
```

Feature	Value
facility_type	0.09
building_class	0.86
energy_star_rating	0.21
days_above_110F	-0.02
floor_area	-0.37
State_Factor	0.49
ELEVATION	-0.60
cooling_degree_days	0.73
may_max_temp	1.05
september_avg_temp	0.61

- The LIME model intercept: 79.40720440515193
- The LIME model prediction: "Prediction_local [94.854192]"
- The original XGBOOST model prediction: "Right: 74.22026"

The local lime prediction =intercept+sum of coefficients

Figure 3. LIME Detailed Result

The bar lengths represent the strength of each feature's impact, while the numbers at the end of the bars indicate the mean SHAP value associated with the features. The predicted value is given in the center, which is the output from the XGBoost model for this instance, and it is compared against the intercept, which represents the base value without any feature influence. The LIME model's intercept and the coefficients for each feature are combined to produce the LIME prediction, which is a locally faithful approximation of the

XGBoost model's complex decision function. This local prediction is based on a linear model that approximates the prediction surface in the vicinity of the instance being explained, aiming to capture the behavior of the complex model as closely as possible for that specific case. Now, while SHAP offers both global and local interpretability, providing insights into feature importance across the entire dataset as well as for individual predictions, it requires calculating SHAP values for each feature by training models across all possible feature combinations. This exhaustive approach provides a very detailed and accurate interpretation but can be computationally intensive and time-consuming, especially for larger datasets with many features.

In contrast, LIME focuses solely on local interpretability. It simplifies the explanation by sampling the vicinity of the instance and using a surrogate linear model to approximate the predictions of the original complex model. This approach is generally faster and less computationally expensive than SHAP, as it avoids the need for global recalculations. LIME's simplicity in providing quick and understandable explanations for individual predictions makes it particularly useful when speed is a concern, or when explanations are needed on-the-fly, for instance, in real-time applications or when many instances need to be explained individually to users or stakeholders.Therefore, while SHAP is comprehensive, LIME's efficiency makes it a pragmatic choice for scenarios where local explanations are sufficient, and computational resources or time are limited. The decision to use LIME over SHAP, or vice versa, will depend on the specific requirements of the explanation task, the size of the dataset, the computational resources available, and the balance between the need for speed versus the need for comprehensive interpretability.

5. CONCLUSION

Throughout this research, we have embarked on a meticulous exploration of building energy usage prediction using advanced machine learning models, with a particular focus on the interpretability of these models through Explainable AI (XAI) techniques. By applying Permutation Importance, SHAP, and LIME methodologies, we have uncovered not only the factors that most significantly influence building energy consumption but also illuminated the decision-making processes of complex predictive models. Our findings have demonstrated that features such as 'Energy Star Rating', 'Facility Type', and 'Floor Area' are paramount in influencing the energy consumption predictions, with weather-related variables also playing a non-negligible role. The Permutation Importance results highlighted the global significance of these features across the dataset, while SHAP values provided both global and local interpretability, offering a deeper dive into how each feature affects individual predictions.

LIME's contribution to our research has been particularly notable in its provision of fast, understandable, local interpretations, emphasizing its utility in scenarios where computational efficiency and speed are crucial. While SHAP's detailed approach to feature importance calculation is invaluable, LIME serves as an agile tool for real-time or on-demand explanations, making it an indispensable asset in the toolkit of data scientists and industry practitioners alike. In conclusion, this study underscores the critical importance of model interpretability in the domain of energy efficiency. The ability to explain model predictions is not merely an academic exercise but a practical necessity that has significant implications for the development of trust in AI systems, the formulation of energy policies, and the implementation of energy conservation measures. As the demand for AI-driven decision-making in energy management continues to grow, the significance of tools like SHAP and LIME will only become more pronounced. As we look to the future, the integration of interpretability into AI systems will likely be a key determinant of their success and acceptance. The methodologies discussed herein will undoubtedly evolve, and new techniques will emerge. Nonetheless, the principles of transparency, accountability, and ethical consideration that have guided this research will remain central to the responsible deployment of AI in energy management and beyond. The path forward is one of continuous improvement, where each step taken towards explainability not only enhances our understanding but also strengthens the foundation for responsible AI.

REFERENCES

[1] Y. Himeur *et al.*, “A survey of recommender systems for energy efficiency in buildings: Principles, challenges and prospects,” *Inf. Fusion*, vol. 72, pp. 1–21, 2021.

[2] S. E. Bibri and J. Krogstie, “Environmentally data-driven smart sustainable cities: Applied innovative solutions for energy efficiency, pollution reduction, and urban metabolism,” *Energy Informatics*, vol. 3, pp. 1–59, 2020.

[3] S. E. Bibri and J. Krogstie, “A novel model for data-driven smart sustainable cities of the future: A strategic roadmap to transformational change in the era of big data,” 2021.

[4] C. W. Chau and E. M. Gerber, “On Hackathons: A Multidisciplinary Literature Review,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–21.

[5] E. Maemura, “Data Here and There: Studying Web Archives Research Infrastructures in Danish and Canadian Settings,” University of Toronto (Canada), 2021.

[6] A. Gaiba, “Demoicratic catalysts: digital technology and institutional change in the Conference on the Future of Europe,” European University Institute, 2022.

[7] V. Masson *et al.*, “City-descriptive input data for urban climate models: Model requirements, data sources and

challenges," *Urban Clim.*, vol. 31, p. 100536, 2020.

[8] O. Bin, T. W. Crawford, J. B. Kruse, and C. E. Landry, "Viewscapes and flood hazard: Coastal housing market response to amenities and risk," *Land Econ.*, vol. 84, no. 3, pp. 434–448, 2008.

[9] R. L. Ciurean *et al.*, "Multi-scale debris flow vulnerability assessment and direct loss estimation of buildings in the Eastern Italian Alps," *Nat. hazards*, vol. 85, pp. 929–957, 2017.

[10] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. fusion*, vol. 58, pp. 82–115, 2020.

[11] V. Chamola, V. Hassija, A. R. Sulthana, D. Ghosh, D. Dhingra, and B. Sikdar, "A review of trustworthy and explainable artificial intelligence (xai)," *IEEE Access*, 2023.

[12] T. Speith, "A review of taxonomies of explainable artificial intelligence (XAI) methods," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 2239–2250.

[13] T. Saroglou, I. A. Meir, T. Theodosiou, and B. Givoni, "Towards energy efficient skyscrapers," *Energy Build.*, vol. 149, pp. 437–449, 2017.

[14] Z. Kang, *Improving Energy Efficiency Performance of Existing Residential Building in Northern China*. Rochester Institute of Technology, 2019.

[15] T. Alves, L. Machado, R. G. de Souza, and P. de Wilde, "Assessing the energy saving potential of an existing high-rise office building stock," *Energy Build.*, vol. 173, pp. 547–561, 2018.

[16] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–42, 2018.

[17] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against machine learning," in *Proceedings of the 2017 ACM on Asia conference on computer and communications security*, 2017, pp. 506–519.

[18] A. J. London, "Artificial intelligence and black-box medical decisions: accuracy versus explainability," *Hastings Cent. Rep.*, vol. 49, no. 1, pp. 15–21, 2019.

[19] J. Borrego-D\'\iaz and J. Galán-Páez, "Explainable Artificial Intelligence in Data Science: From Foundational Issues Towards Socio-technical Considerations," *Minds Mach.*, vol. 32, no. 3, pp. 485–531, 2022.

[20] D. Diepgrond, "Can prediction explanations be trusted? On the evaluation of interpretable machine learning methods," 2020.

[21] V. Palmisano, "Responsible Artificial Intelligence for Critical Decision-Making Support: A Healthcare Scenario," Politecnico di Torino, 2022.

[22] S. Hariharan, R. R. Rejimol Robinson, R. R. Prasad, C. Thomas, and N. Balakrishnan, "XAI for intrusion detection system: comparing explanations based on global and local scope," *J. Comput. Virol. Hacking Tech.*, vol. 19, no. 2, pp. 217–239, 2023.

[23] G. Fidel, R. Bitton, and A. Shabtai, "When explainability meets adversarial learning: Detecting adversarial examples using shap signatures," in *2020 international joint conference on neural networks (IJCNN)*, 2020, pp. 1–8.

[24] H. T. T. Nguyen, H. Q. Cao, K. V. T. Nguyen, and N. D. K. Pham, "Evaluation of explainable artificial intelligence: Shap, lime, and cam," in *Proceedings of the FPT AI Conference*, 2021, pp. 1–6.

[25] Y. Wang and Y. Guo, "Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost," *China Commun.*, vol. 17, no. 3, pp. 205–221, 2020.

[26] K. Liu, X. Xu, R. Zhang, L. Kong, W. Wang, and W. Deng, "Impact of urban form on building energy consumption and solar energy potential: A case study of residential blocks in Jianhu, China," *Energy Build.*, vol. 280, p. 112727, 2023.

[27] J. Alzubi, A. Nayyar, and A. Kumar, "Machine learning from theory to algorithms: an overview," in *Journal of physics: conference series*, 2018, vol. 1142, p. 12012.

[28] V. Moustaka, Z. Theodosiou, A. Vakali, A. Kounoudes, and L. G. Anthopoulos, "Enhancing social networking in smart cities: Privacy and security borderlines," Technol. Forecast. Soc. Change, vol. 142, pp. 285–300, 2019.

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.