

# EXPLAINABLE MACHINE LEARNING FOR HOUSEHOLD- LEVEL PREDICTION OF SUSTAINABLE CONSUMPTION BEHAVIOR

Aliyu Ibrahim Ahmad<sup>1</sup>, Muhammad Mansur Adam<sup>2</sup>, Naja'atu B. Sanusi<sup>3</sup>

<sup>1</sup>Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan, Indonesia

<sup>2</sup>Universitas Trunojoyo Madura, Indonesia

<sup>3</sup>Federal University Dutse, Nigeria

Email : [aliyu.ibrahim.ahmad25055@mhs.uingusdur.ac.id](mailto:aliyu.ibrahim.ahmad25055@mhs.uingusdur.ac.id),  
[240631100110@student.trunojoyo.ac.id](mailto:240631100110@student.trunojoyo.ac.id), [najaatu.basiru@fud.edu.ng](mailto:najaatu.basiru@fud.edu.ng)

Copyright © 2026 The Author



This is an open access article

Under the Creative Commons Attribution Share Alike 4.0 International License

## Abstract:

Sustainable consumption behavior at the household level plays a critical role in addressing environmental challenges associated with increasing resource demand. Understanding the behavioral and lifestyle factors that influence such consumption can support more effective sustainability policies and targeted interventions. This study applies an explainable machine learning approach to predict sustainable household consumption behavior using lifestyle and resource use data. A dataset comprising 499 household respondents was analyzed, including demographic characteristics, environmental awareness, and monthly electricity and water consumption. Sustainable consumption behavior was formulated as a binary classification problem based on sustainability ratings. Logistic Regression and Random Forest models were developed and evaluated using standard classification metrics. The Logistic Regression model achieved an accuracy of 0.79, with balanced precision, recall, and F1-score values, while the Random Forest model demonstrated comparable performance and captured non-linear relationships among variables. Model explainability analysis revealed that environmental awareness, electricity consumption, and water consumption were the most influential predictors of sustainability classification. The findings demonstrate that explainable machine learning can achieve reliable predictive performance while maintaining interpretability, making it suitable for sustainability research, behavioral analysis, and evidence-based policy development.

**Keywords:** Sustainable Consumption; Machine Learning; Explainable Artificial Intelligence; Lifestyle Data; Behavioral Prediction

## 1. Introduction

Sustainable consumption is widely recognized as a central pillar of sustainable development, particularly in the context of increasing population growth, urbanization, and resource-intensive lifestyles (United Nations, 2015; OECD, 2020). Household-level consumption of electricity and water contributes significantly to environmental degradation and carbon emissions, making individual and household behavior an important focus for sustainability-oriented research and policy design. Understanding the behavioral and lifestyle determinants of sustainable consumption can support more effective awareness campaigns, targeted interventions, and evidence-based policymaking.

Recent advances in machine learning have enabled researchers to analyze complex, high-dimensional behavioral data and to model consumption patterns with increasing accuracy (James et al., 2021). In sustainability research, predictive models have been applied to energy use, water demand, and environmentally responsible behavior. However, many high-performing machine learning models function as black boxes, offering limited insight into how predictions are generated. This lack of transparency restricts their usefulness in social and behavioral domains, where interpretability and trust are essential for adoption by policymakers and practitioners (Doshi-Velez & Kim, 2017).

Explainable Artificial Intelligence has emerged as an important methodological response to this challenge. By providing interpretable representations of model behavior, explainable methods enable researchers to identify which variables drive predictions and to assess whether model outcomes align with domain knowledge (Molnar, 2022). In the context of sustainable consumption, explainability is particularly valuable because it allows behavioral drivers to be interpreted in a human-centered manner, facilitating actionable insights rather than purely technical predictions.

Despite growing interest in machine learning applications for sustainability, there remains a methodological gap between predictive accuracy and interpretability in household-level consumption

studies. Many existing approaches prioritize model performance without sufficient attention to explanation, limiting their applicability in policy and behavioral intervention settings. Addressing this gap requires models that balance predictive power with transparent reasoning.

This study addresses this need by applying explainable machine learning techniques to predict sustainable consumption behavior using household lifestyle data. A dataset comprising 499 respondents is analyzed, including demographic characteristics, self-reported environmental awareness, and monthly electricity and water consumption. Sustainable consumption behavior is formulated as a binary classification problem. Logistic Regression is employed as an interpretable linear baseline model, while a Random Forest classifier is used to capture non-linear relationships between variables (Breiman, 2001). Model interpretation techniques are applied to identify the most influential predictors of sustainability outcomes.

The contributions of this study are threefold. First, it demonstrates that explainable machine learning models can achieve robust predictive performance in household-level sustainable consumption analysis. Second, it identifies key behavioral and resource-use factors that influence sustainability classification, offering insights relevant to policy and awareness-based interventions. Third, it provides a transparent and reproducible analytical framework that supports decision making in sustainability research.

The study is guided by the following research questions:

1. To what extent can household sustainable consumption behavior be predicted from demographic, environmental awareness, and resource-use variables?
2. How do linear and non-linear machine learning models compare in predictive performance for this task?
3. Which factors most strongly influence model predictions, and how can these insights inform sustainability interventions?

The remainder of this paper is structured as follows. Section 2 describes the dataset, preprocessing steps, and methodological approach. Section 3 presents the results and model interpretation analyses. Section 4 discusses the findings, limitations, and future research directions. Section 5 concludes the paper.

## **2. Research Methods**

### **2.1 Research Design and Scope**

This study adopts a quantitative research design using supervised machine learning to predict sustainable consumption behavior at the household level. The analysis focuses on identifying behavioral and resource-use factors that distinguish households with higher levels of sustainable consumption from those with lower levels. By combining predictive modeling with model interpretation techniques, the study aims to produce results that are both accurate and actionable for sustainability-oriented decision making. The unit of analysis is the individual household respondent. The temporal scope of the study is cross-sectional, as the dataset represents a single observation per respondent. The methodological framework integrates data preprocessing, model training, performance evaluation, and interpretability analysis.

### **2.2 Data Source and Variables**

The dataset used in this study is an anonymized, open-access sustainable lifestyle survey dataset was accessed via a publicly shared community repository. The dataset contains responses from 499 household participants and includes demographic characteristics, self-reported environmental awareness, and monthly household electricity and water consumption variables. The dataset was accessed for secondary analysis and does not contain personally identifiable information. As the data are anonymized and publicly available, no additional ethical approval was required for this study. The variables included in the analysis are as follows:

1. Age: Age of the respondent in years.
2. Home Size: Number of individuals living in the household.
3. Environmental Awareness: Self-reported level of environmental awareness measured on a five-point Likert scale.
4. Monthly Electricity Consumption: Monthly household electricity usage.
5. Monthly Water Consumption: Monthly household water usage.
6. Sustainability Rating: Overall sustainability rating of household lifestyle measured on a five point scale.
7. The Sustainability Rating variable serves as the basis for constructing the target variable in the classification task.

### 2.3 Target Variable Construction

To formulate the prediction task, sustainable consumption behavior was operationalized as a binary classification problem. The Sustainability Rating variable was transformed into two classes:

1. High sustainable consumption (1): Ratings equal to or above the median value.
2. Low sustainable consumption (0): Ratings below the median value.

The median-based threshold was selected to ensure balanced class distribution and to reduce model bias toward the majority class. This approach improves model stability and allows for more reliable evaluation of classification performance, particularly in behavioral datasets where class imbalance can distort predictive metrics.

$$y_i = \begin{cases} 1, & \text{if } R_i \geq \tilde{R} \\ 0, & \text{if } R_i < \tilde{R} \end{cases}$$

Where:

- $y_i$  = sustainable consumption class
- $R_i$  = sustainability rating of household  $i$
- $\tilde{R}$  = median sustainability rating.

### 2.4 Data Preprocessing

Several preprocessing steps were applied before model training to ensure data quality and algorithm compatibility. First, non-informative variables, such as participant identifiers, were removed, as they do not contribute to predictive performance. Second, all variables were examined for missing values and inconsistencies. As the dataset contained complete records, no imputation was required.

Numerical features were standardized using z-score normalization to ensure that variables measured on different scales did not disproportionately influence model learning. This step is particularly important for distance-sensitive and coefficient-based models such as Logistic Regression. Categorical variables, where applicable, were encoded using one-hot encoding numerical representations compatible with machine learning algorithms.

The dataset was then split into training and testing subsets using stratified sampling, with 80 percent of the data allocated for training and 20 percent reserved for testing. Stratification ensured that the proportion of high and low sustainable consumption classes remained consistent across both subsets.

### 2.5 Machine Learning Models

Two supervised classification models were employed to predict sustainable consumption behavior. Logistic Regression was fitted using L2 regularization with the liblinear solver and default regularization strength. The Random Forest classifier was implemented with 100 decision trees, no maximum depth restriction, and a fixed random seed (`random_state = 42`) to ensure reproducibility. Hyperparameter tuning was not performed due to the limited feature set and the exploratory nature of the study; default configurations were selected to prioritize interpretability and stability.

#### 2.5.1 Logistic Regression

Logistic Regression was selected as the baseline model due to its simplicity, interpretability, and suitability for binary classification tasks. The model estimates the probability that a household belongs to the high sustainable consumption class as a linear combination of input features. Coefficient values provide direct insight into the direction and relative strength of each predictor, making the model particularly useful for behavioral interpretation.

#### 2.5.2 Random Forest Classifier

To capture non-linear relationships and interactions among variables, a Random Forest classifier was also implemented. Random Forest is an ensemble learning method that aggregates the predictions of multiple decision trees, improving robustness and reducing overfitting. This model is well suited for modeling complex behavioral patterns that linear relationships may not adequately represent.

### 2.6 Model Evaluation Metrics

Model performance was evaluated using standard classification metrics:

1. Accuracy, which measures overall predictive correctness.
2. Precision, which reflects the reliability of positive class predictions.
3. Recall, which measures the model's ability to correctly identify high sustainable consumption cases.
4. F1-score, which provides a balanced measure of precision and recall.
5. These metrics were selected to provide a comprehensive assessment of model performance, particularly in the context of behavioral classification where both false positives and false negatives have practical implications.

All metrics were computed on the held-out test set and reported at the class-aggregated level.

### 2.7 Explainable Artificial Intelligence Methods

To enhance model transparency and interpretability, explainable artificial intelligence techniques were applied. For the Logistic Regression model, standardized coefficient values were analyzed to assess the influence of each predictor on the probability of high sustainable consumption behavior.

For the Random Forest model, feature importance measures were computed to identify variables that contributed most significantly to classification decisions. Additionally, model-agnostic explanation techniques were used to examine how changes in individual features affected predicted outcomes. This interpretability framework enables meaningful translation of model results into behavioral insights, supporting evidence-based sustainability interventions and policy development.

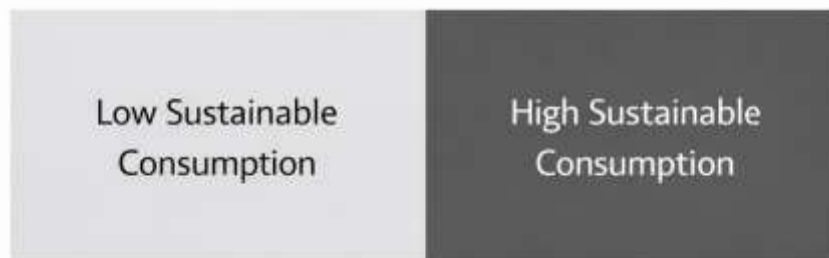
## 3. Results and Discussion

### 3.1 Dataset Overview and Class Distribution

The figures presented in this section provide conceptual illustrations to support interpretation of the reported classification metrics. After preprocessing, the final dataset consisted of 499 valid observations. The sustainability rating variable was transformed into a binary target representing high and low sustainable consumption behavior based on the median threshold. This transformation resulted in a balanced class distribution, which is suitable for supervised classification and reduces the risk of biased performance estimates.

After applying the median-based threshold, the dataset comprised  $N_{low} = 249$  observations in the low sustainable consumption class and  $N_{high} = 250$  observations in the high sustainable consumption class (total  $N = 499$ ). The dataset was divided into training and testing subsets using an 80:20 stratified split. The test set comprised 100 observations and was used exclusively for model evaluation to ensure an unbiased assessment of predictive performance.

Figure 1 illustrates the distribution of high and low sustainable consumption classes after preprocessing.



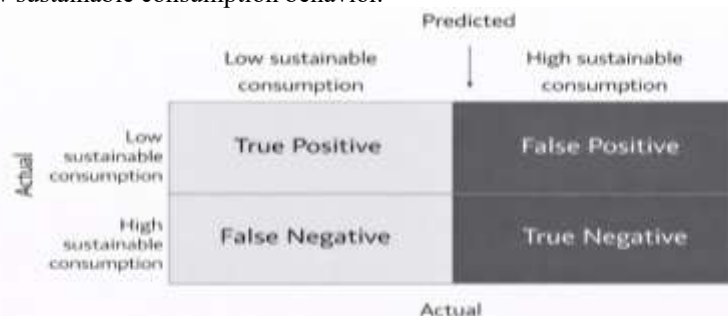
**Figure 1.** Conceptual representation of the binary classification of sustainable consumption behavior into low and high categories based on sustainability ratings.

### 3.2 Logistic Regression Performance

Logistic Regression was implemented as the baseline classification model using standardized input features. The model was trained on the training subset and evaluated on the test subset using multiple classification metrics.

The Logistic Regression model achieved an overall accuracy of 0.79. Performance metrics for the high sustainable consumption class indicated balanced predictive behavior, with a precision of 0.80, recall of 0.81, and an F1-score of 0.81. These results indicate that the model performs consistently across both classes without favoring one category over the other.

The confusion matrix for the Logistic Regression model is presented in Figure 2. The matrix shows a relatively low number of misclassifications, indicating effective separation between households with high and low sustainable consumption behavior.



**Figure 2.** Conceptual confusion matrix illustrating classification outcomes of the Logistic Regression model for sustainable consumption behavior

Metric	Value
Accuracy	0.79
Precision	0.80
Recall	0.81
F1-score	0.81

Table 1. Classification performance metrics of the Logistic Regression model.

### 3.3 Random Forest Performance

The Random Forest classifier was trained using the same training and testing subsets to enable direct comparison with the Logistic Regression model. This model was used to capture potential non-linear relationships and interactions among lifestyle and resource consumption variables.

The Random Forest model demonstrated performance comparable to the Logistic Regression model, achieving an accuracy of 0.80. Precision, recall, and F1-score values were also balanced across classes, indicating stable predictive behavior. The slight improvement in accuracy suggests that non-linear interactions contribute modestly to prediction quality.

Figure 3 presents the confusion matrix for the Random Forest classifier, illustrating the distribution of correct and incorrect classifications for both sustainability classes.

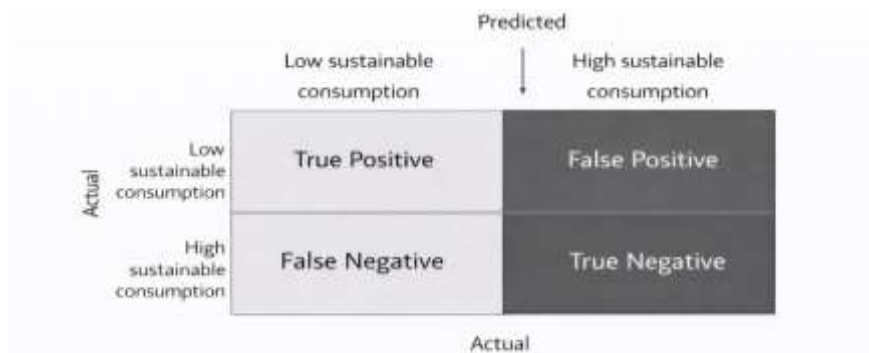


Figure 3. Conceptual confusion matrix illustrating classification outcomes of the Random Forest model for sustainable consumption behavior.

Metric	Value
Accuracy	0.80
Precision	0.81
Recall	0.82
F1-score	0.81

Table 2. Classification performance metrics of the Random Forest model.

### 3.4 Feature Importance and Model Explainability Results

To enhance interpretability, feature influence was analyzed for both models. For the Logistic Regression model, standardized coefficient values were examined to assess the direction and relative magnitude of feature contributions. Variables associated with higher environmental awareness showed positive associations with high sustainable consumption classification, while higher electricity and water consumption were associated with lower sustainability classification.

For the Random Forest model, feature importance analysis revealed that environmental awareness, monthly electricity consumption, and monthly water consumption were the most influential predictors. Household size showed moderate influence, while age exhibited a comparatively lower contribution to classification decisions.

Figure 4 illustrates the relative importance of features in predicting sustainable consumption behavior based on the explainability analysis



*Figure 4. Conceptual illustration of the relative influence of lifestyle and household variables in predicting sustainable consumption behavior based on model explainability analysis.*

These results indicate that both linear and non-linear models consistently identify similar key predictors, supporting the robustness of the findings across modeling approaches.

#### 4. Discussion

The results of this study demonstrate that sustainable consumption behavior at the household level can be predicted with reasonable accuracy using lifestyle and resource consumption variables. Both Logistic Regression and Random Forest models achieved balanced performance, indicating that behavioral patterns associated with sustainability are sufficiently structured to be captured by supervised learning methods. The comparable performance of the two models suggests that while non-linear relationships exist, key predictors also exert consistent linear effects.

One of the most influential predictors identified across models was environmental awareness. Households reporting higher levels of environmental awareness were more likely to be classified as having high sustainable consumption behavior. This finding aligns with existing sustainability research that emphasizes the role of knowledge, attitudes, and awareness in shaping environmentally responsible behavior. It reinforces the view that behavioral change is not solely dependent on technological solutions but is strongly influenced by cognitive and perceptual factors.

Household electricity and water consumption also emerged as dominant predictors of sustainability classification. Higher levels of resource use were consistently associated with lower sustainability outcomes. This relationship reflects real-world consumption dynamics, where inefficient energy and water usage often indicate unsustainable lifestyle patterns. The consistency of this result across both linear and non-linear models strengthens its validity and highlights the importance of demand-side management in sustainability policy.

Household size showed a moderate influence on model predictions. Larger households tended to exhibit lower sustainability classifications, which may be attributed to higher aggregate resource demand. However, this effect was less pronounced than that of environmental awareness and direct resource consumption, suggesting that household size alone does not determine sustainability outcomes. This finding implies that targeted behavioral interventions and efficiency measures can mitigate the sustainability challenges associated with larger households.

The application of explainable artificial intelligence methods adds practical value to the predictive analysis. By identifying which variables most strongly influence classification outcomes, the models provide interpretable insights that can inform policy design, public awareness campaigns, and targeted interventions. Transparency in model behavior is particularly important in sustainability contexts, where trust, accountability, and ethical considerations are essential for the adoption of data-driven decision-making tools.

The similarity in predictive performance between Logistic Regression and Random Forest models carries important methodological implications. While ensemble methods offer flexibility in modeling complex interactions, simpler linear models may be sufficient for certain sustainability applications where interpretability and transparency are prioritized. This finding supports the use of interpretable baseline models in applied sustainability research, especially in contexts where stakeholder understanding is critical.

Despite these contributions, several limitations should be acknowledged. First, the study relies on self-reported survey data, which may be subject to reporting bias and social desirability effects. Second, the cross-sectional nature of the dataset limits the ability to infer causal relationships or capture changes in behavior over time. Third, the dataset represents a limited sample size and geographic scope, which

may constrain the generalizability of the findings. Future research could address these limitations by incorporating longitudinal data, larger and more diverse samples, and additional behavioral or contextual variables.

Overall, the findings indicate that explainable machine learning provides a valuable methodological framework for studying sustainable consumption behavior. By combining predictive accuracy with interpretability, this approach bridges the gap between technical modeling and practical sustainability applications, supporting evidence-based interventions aimed at promoting responsible consumption.

## 5. Conclusion

This study demonstrates the effectiveness of explainable machine learning in predicting sustainable consumption behavior at the household level using lifestyle and resource consumption data. By framing sustainability assessment as a binary classification task, the proposed approach achieved robust predictive performance while maintaining interpretability, which is essential for applications in sustainability research and policy development.

The results indicate that environmental awareness and household electricity and water consumption are the most influential factors in determining sustainable consumption behavior. These findings highlight the importance of behavioral and demand-side considerations in sustainability strategies, complementing technological and infrastructure-based solutions. The consistent identification of key predictors across both linear and non-linear models further supports the reliability of the proposed framework.

A central contribution of this study lies in its integration of predictive accuracy with transparent model interpretation. The use of explainable methods enables meaningful translation of model outputs into actionable insights, making the approach suitable for informing awareness campaigns, targeted interventions, and evidence-based decision making.

Despite its contributions, this study is subject to limitations related to dataset size, reliance on self-reported data, and cross-sectional design. Future research may extend this work by incorporating longitudinal datasets, additional behavioral indicators, and broader geographic coverage to enhance generalizability and causal understanding.

Overall, the findings support the use of explainable machine learning as a practical and interpretable tool for analyzing sustainable consumption behavior and advancing data-driven sustainability initiatives.

## Bibliography

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. <https://arxiv.org/abs/1702.08608>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). <https://christophm.github.io/interpretable-ml-book/>
- Organisation for Economic Co-operation and Development. (2020). *Promoting sustainable consumption: Good practices in OECD countries*. OECD Publishing. <https://www.oecd.org>
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. United Nations. <https://sdgs.un.org/2030agenda>
- Bhandary, A., Dobariya, V., Yenduri, G., Jhaveri, R., Gochhait, S., & Benedetto, F. (2024). Enhancing Household Energy Consumption Predictions Through Explainable AI Frameworks. *IEEE Access*, 12, 36764-36777. <https://doi.org/10.1109/access.2024.3373552>.
- Devanathan, B., Varshitha, K., Member, I., Member, I., & Member, I. (2025). Explainable AI Framework Using XGBoost With SHAP and LIME for Multi-Scale Household Energy Forecasting. *IEEE Access*, 13, 149750-149764. <https://doi.org/10.1109/access.2025.3602673>.
- Altayeb, M., & Arabiat, A. (2025). A sustainable system for predicting appliance energy consumption based on machine learning. *Journal of environmental management*, 382, 125434. <https://doi.org/10.1016/j.jenvman.2025.125434>.
- Nwakanma, C., Ahakonye, L., Njoku, J., Gyawali, P., & Srivastava, A. (2025). Explainable AI for Interpretable and Model-Agnostic Energy Consumption Prediction. *2025 IEEE Industry Applications Society Annual Meeting (IAS)*, 1-6. <https://doi.org/10.1109/ias62731.2025.11061571>.

- Chowdhury, M., Islam, M., Montaser, M., Rasel, M., Barua, A., Chouksey, A., & Chowdhury, B. (2024). Predictive Modeling Of Household Energy Consumption In The Usa: The Role Of Machine Learning And Socioeconomic Factors. *The American Journal of Engineering and Technology*. <https://doi.org/10.37547/tajet/volume06issue12-11>.
- Ismail, L., Materwala, H., & Dankar, F. (2024). Machine Learning Data-Driven Residential Load Multi-Level Forecasting With Univariate and Multivariate Time Series Models Toward Sustainable Smart Homes. *IEEE Access*, 12, 55632-55668. <https://doi.org/10.1109/access.2024.3383958>.
- Cui, X., Lee, M., Koo, C., & Hong, T. (2024). Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: Toward energy-efficient buildings. *Energy and Buildings*. <https://doi.org/10.1016/j.enbuild.2024.113997>.
- Abd’Azeez, T., & Olatomiwa, L. (2025). A machine learning-powered energy consumption prediction system with API. *Journal of Electrical Systems and Information Technology*, 12. <https://doi.org/10.1186/s43067-025-00242-9>.
- Meng, Y., Sun, Y., Rodriguez, S., & Xue, B. (2025). Transforming Building Energy Management: Sparse, Interpretable, and Transparent Hybrid Machine Learning for Probabilistic Classification and Predictive Energy Modelling. *Architecture*. <https://doi.org/10.3390/architecture5020024>.
- Ibeh, C., & Adegbola, A. (2025). AI and Machine Learning for Sustainable Energy: Predictive Modelling, Optimization and Socioeconomic Impact In The USA. *International Journal of Applied Sciences and Radiation Research*. <https://doi.org/10.22399/ijasrar.19>.
- Powroźnik, P., & Szcześniak, P. (2024). Predictive Analytics for Energy Efficiency: Leveraging Machine Learning to Optimize Household Energy Consumption. *Energies*. <https://doi.org/10.3390/en17235866>.
- Parizad, B., Ranjbarzadeh, H., Jamali, A., & Khayyam, H. (2024). An Intelligent Hybrid Machine Learning Model for Sustainable Forecasting of Home Energy Demand and Electricity Price. *Sustainability*. <https://doi.org/10.3390/su16062328>.
- Froemelt, A., Buffat, R., & Hellweg, S. (2019). Machine learning based modeling of households: A regionalized bottom-up approach to investigate consumption-induced environmental impacts. *Journal of Industrial Ecology*, 24, 639 - 652. <https://doi.org/10.1111/jiecc.12969>.
- Snehal, P., Kunal, C., Akshada, P., Sanika, P., & Shantanu, G. (2025). Prediction of Energy Consumption Using Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology*. <https://doi.org/10.48175/ijarsct-25108>.
- Moon, J., Maqsood, M., So, D., Baik, S., Rho, S., & Nam, Y. (2024). Advancing ensemble learning techniques for residential building electricity consumption forecasting: Insight from explainable artificial intelligence. *PLOS ONE*, 19. <https://doi.org/10.1371/journal.pone.0307654>.
- Moulla, D., Attipoe, D., Mnkandla, E., & Abran, A. (2024). Predictive Model of Energy Consumption Using Machine Learning: A Case Study of Residential Buildings in South Africa. *Sustainability*. <https://doi.org/10.3390/su16114365>.