Explainable AI for Energy Prediction and Anomaly Detection in Smart Energy Buildings

Hardik Prabhu
Department of Computing and Data Sciences, FLAME University
Pune, India
hardik.prabhu@gmail.com

Jayaraman Valdi
Department of Computing and Data Sciences, FLAME University
Pune, India
jayaraman.vk@flame.edu.in

Pandarasamy Arjunan
Robert Bosch Centre for Cyber Physical Systems
Indian Institute of Science
Bangalore, India
samy@iisc.ac.in

ABSTRACT

In recent years, the advancement of Artificial Intelligence (AI) and Advanced Metering Infrastructure (AMI) has led to the development of data-driven methods for energy prediction and anomaly detection. These methods provide automated decision support to building operators in managing and preventing energy loss. Despite the advantages of having sophisticated data-driven models, one major drawback is their lack of transparency, which limits their widespread use. The paper explores the use of the SHapely Additive exPlanations (SHAP), an explainable AI algorithm, to enhance transparency in energy prediction and anomaly detection models. Energy prediction is treated as a regression task, while anomaly detection as a binary classification. The study employs LightGBM models for both anomaly detection and energy prediction, which are tested on a large dataset containing hourly smart metering data from over 200 real buildings. The energy prediction model achieves an R^2 score of 0.975, while the anomaly detection model obtains an AUC-ROC score of 0.942. These models are augmented with SHAP value-based visualizations, which provide both local and global explanations of these models, offering valuable insights into the factors influencing their predictions. Additionally, the present study introduces a framework that seamlessly integrates feature transformations within the model, while SHAP operates on the interpretable feature space, enhancing the explanations provided by SHAP values.

CCS CONCEPTS

• **Computing methodologies → Machine learning algorithms**; **Boosting**; **Anomaly detection**; **Supervised learning by regression**.

KEYWORDS

XAI, Energy Prediction, Anomaly Detection, LightGBM

ACM Reference Format:
Hardik Prabhu, Jayaraman Valdi, and Pandarasamy Arjunan. 2023. Explainable AI for Energy Prediction and Anomaly Detection in Smart Energy Buildings. In *The 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23), November 15–16, 2023, Istanbul, Turkey.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3600100.3626638

1 INTRODUCTION

To meet our sustainability goals, reducing energy consumption in the building sector is essential. A promising solution is the widespread use of energy metering infrastructure, now adopted globally, leading to a wealth of building energy data. This data offers opportunities for data-driven energy forecasting, enabling effective energy usage prediction and optimization by building managers [6]. Recognizing the impact of faults is vital, as in commercial buildings, poorly maintained hardware and operational issues can waste 15 to 30 % of energy consumption [2]. Both anomaly detection and energy prediction are key for efficient energy management.

When addressing energy prediction and anomaly detection at a large scale, it is beneficial to leverage complex data-driven models. However, these models often operate as black-boxes due to the challenges in interpreting their internal workings. Lack of interpretability hinders building operators' use of these models. Our goal is to make black-box models more transparent, helping operators understand, trust, and derive insights from their predictions.

Recently several explainable AI-based algorithms have been proposed to unbox the black-box models in the smart grid domain and other energy-related domains [4]. To improve black-box model interpretability, we leverage the widely used SHapely Additive exPlanations (SHAP) [3]. We also introduce a framework that seamlessly integrates feature transformations within the model. This enhances the interpretability of SHAP explanations by operating on the human-interpretable feature space, while the model operates on the transformed space. We evaluate the proposed method on a large dataset containing 1,749,494 readings and show that the proposed method helps enhance model transparency.

2 METHODOLOGY

2.1 Dataset

Building Data Genome 2 (BDG2) is an open-source data set of hourly energy meter readings from 1,636 buildings. In 2019, ASHRAE organized the *Great Energy Predictor III* competition on Kaggle, using a subset of BDG2 [5]. Later in 2022, the Large-scale Energy Anomaly Detection (LEAD) competition was hosted on Kaggle

BuildSys '23, November 15–16, 2023, Istanbul, Turkey

ACM ISBN 979-8-4007-0230-3/23/11…$15.00
https://doi.org/10.1145/3600100.3626638

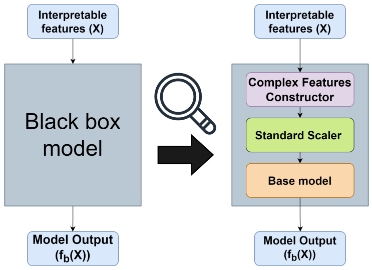

Figure 1: Black-box model in detail

Table 1: Interpretable features extracted from the dataset.

Type	Features
Temporal	hour, weekday, month, is holiday
Weather	air temperature, cloud coverage, dew temperature, precip depth 1 hr, sea level pressure, wind direction, wind speed
Meta	building id, site id, square feet, year built, floorcount

with the aim of detecting abnormal energy consumption and open-sourced annotated meter readings from 400 buildings [1]. Winning solutions in both competitions employed LightGBM and advanced feature engineering. We utilize a subset of the BDG2 and LEAD datasets, which includes hourly meter readings over 12 months from 200 buildings with point-wise anomaly labels.

2.2 Data Pre-processing

For the energy prediction task, we start with the features shown in Table 1. For the anomaly detection task, we create a new dataset by including meter readings (omitted in the prediction task) and incorporate value-change features, inspired by the winning solution in the LEAD competition. Value-change in meter readings, which computes the value change in the form of difference and ratio, are created. Different shift steps are considered to capture changes over various time ranges. The datasets are then split into train-test based on the months, The first 8 months of data is considered as training and the rest is considered as test.

Anomalies comprise less than 3% of the training data, indicating a notable class imbalance. To balance this, we downsampled the normal data by randomly selecting an equal number of normal and anomaly points for anomaly detection training. No downsampling was done for energy prediction.

2.3 Explaining Black-box Models with SHAP

LightGBM is a popular open-source model known for its efficient gradient-boosting tree algorithm. It's versatile, and suitable for tasks like classification and regression on large datasets, making it a preferred choice in both competitions. The model's accuracy can be attributed to its ensemble of decision trees. However, as the number of trees increases, it becomes difficult to understand how the model combines these trees to make individual predictions, essentially rendering it a black-box.

To understand how LightGBM arrives at individual predictions, we use SHAP values, derived from game theory's Shapley values. These values fairly allocate contributions of individual features to a model's prediction. For a given instance x, the additivity property of SHAP represents the prediction $f(x)$ as the following.

$$f(x) = E[f(x)] + \sum_{i=1}^{n} \phi_i \tag{1}$$

Where ϕ_i denotes the feature contribution of the i^{th} feature which is also called its SHAP value.

2.4 Input Feature Transformation and Interpretability

2.4.1 Interpretable Feature Space. Interpretable feature space refers to a representation of the data where the features have a clear meaning. It means that the features in the dataset are designed or selected in a way that can be easily interpreted and understood by humans.

2.4.2 Model Feature Space. In the model feature space, the features are engineered and transformed to optimize the performance of the model. This may involve creating new features from existing ones and applying mathematical transformations, such as standard scaling. For example, during the competitions, it was observed that extracting cyclic coordinates for temporal features proved to be beneficial. For instance, certain features like the weekday exhibit periodicity. Equation 2 shows how to extract cyclic coordinate features, from the original feature where n is the number of unique values the feature takes.

$$feat_x(i), feat_y(i) = cos(2\pi i/n), sin(2\pi i/n) \tag{2}$$

The model feature space obtained after feature engineering and scaling is less human-interpretable. The SHAP framework offers to explain a black-box model by quantifying the impact of individual features on the model's prediction. To enhance interpretability, it is crucial to apply the SHAP framework specifically to the interpretable feature space. In order to restrict SHAP analysis to only the interpretable space, the transformation from the interpretable space to the model space is done within the model. Let f_b be the choice of the base black-box model, which operates on the transformed space. x is the instance in the interpretable space. The model $f = f_b(transform(x))$ along with the data in the interpretable space is provided to the SHAP framework to generate explanations. Refer to Figure 1 for an illustration.

3 EXPERIMENTS AND RESULTS

3.1 Black-Box Model Evaluation

We present the evaluation of our black-box models, comparing them with baseline interpretable models. Linear models such as Linear/Logistic regression are widely utilized for various prediction/binary classification tasks in diverse fields owing to their interpretability and transparency. For the energy prediction model, the metrics selected to evaluate its performance are R-squared (R^2) and Mean Absolute Error (MAE). For the energy anomaly classifier model, the metric chosen to evaluate its performance is the Area Under the Receiver Operating Characteristic curve (AUC-ROC). Our experiments highlight the superior performance of the black-box models compared to the interpretable baseline models in both energy prediction and anomaly detection tasks. These results are summarized in Table 2.

Table 2: Performance comparison of models

Algorithm	Energy prediction		Anomaly detection
	R^2	MAE	AUC-ROC
LightGBM	0.975	28.201	0.942
Linear Model	0.307	144.991	0.628

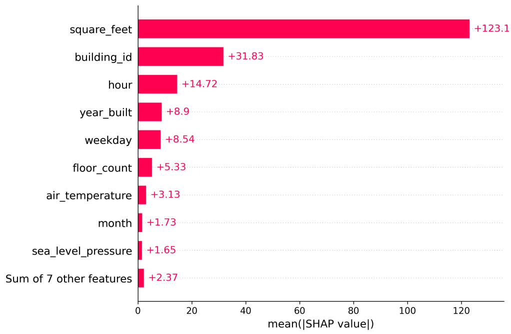

Figure 2: Bar plot for global feature importance

3.2 SHAP Explanations

A LightGBM regression model is trained to predict energy consumption. To gain a global understanding of the model, we select a large data sample and construct a matrix of SHAP values. Each column represents SHAP feature values, and each row corresponds to a specific data point. By examining the mean absolute SHAP values across all data points, we can identify which features significantly impact the model's predictions. Figure 2 presents the global feature importance. It shows the overall influence of a particular feature on the model in order to make predictions. Among the features, square_feet, building_id, and hour emerge as the most influential for the model.

In order to explain the behaviour of this model for a particular instance, the SHAP values are calculated and presented in the form of a waterfall plot. It illustrates the contribution of each feature to the prediction output of the model. Figure 3 depicts a waterfall plot for an arbitrary instance. The base prediction value $(E[f(x)])$ for the model is 140.4. We observe that the size of the building has a significantly larger positive impact on the prediction. Additionally, other factors such as the time of the day and weekday also contribute positively to driving the prediction higher.

Interesting SHAP value patterns emerge with varying feature values and can be analyzed using a scatter plot, where each point represents an instance, plotted with feature value on the x-axis and SHAP value on the y-axis. Figure 4 displays this for the feature square_feet. As its value rises, its positive contribution to the prediction grows, suggesting larger building areas in the dataset may push predictions higher.

To enhance the informative nature of the scatterplot, one option is to assign colours to each point based on the values of another feature at those specific points. This is particularly useful when considering the varying effects of a specific value of a feature, such as a weekday, on different building sizes. Figure 5 shows the dependency plot for feature weekday. The dependency plot for weekday reveals that the weekends (5 and 6) have a negative impact on building energy use while working days have a positive impact on energy use. A similar trend is observed for the feature hour indicating that during non-working hours, there are instances where the hour has a negative impact on the meter reading. Upon examining the corresponding values of the feature square_feet, it becomes apparent that the impact is more pronounced in buildings with larger sizes.

A LightGBM binary classifier is trained for anomaly detection. The model provides a probability score indicating the likelihood of an instance being anomalous. Figure 6 shows the global feature

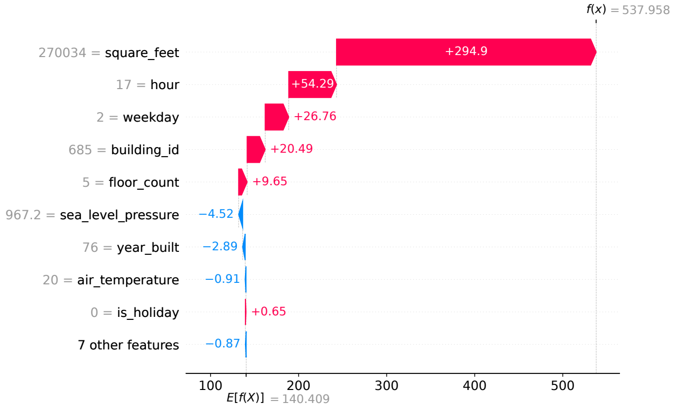

Figure 3: Waterfall plot for a specific model prediction

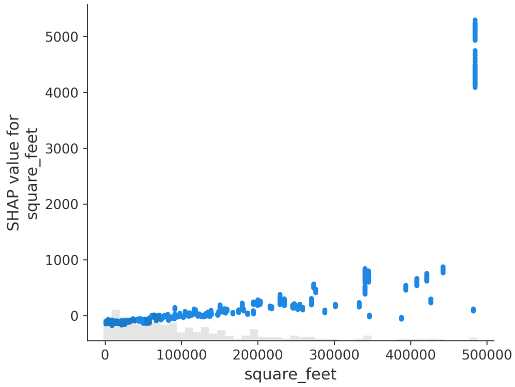

Figure 4: Scatter plot for the feature square_feat

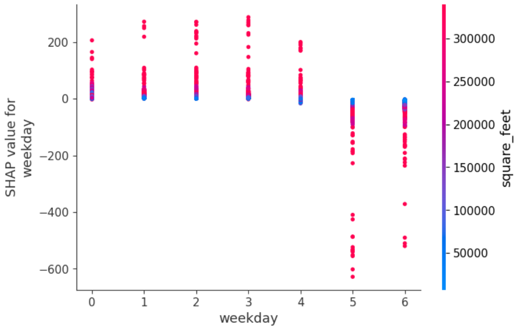

Figure 5: Dependency plot for the feature weekday

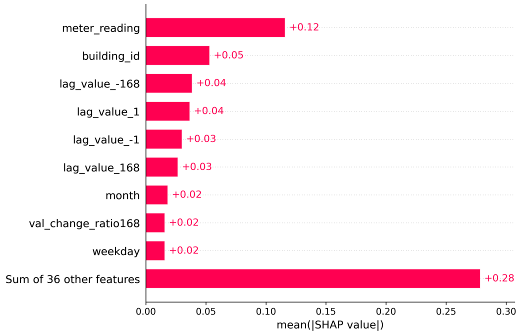

Figure 6: Global feature importance for anomaly detection

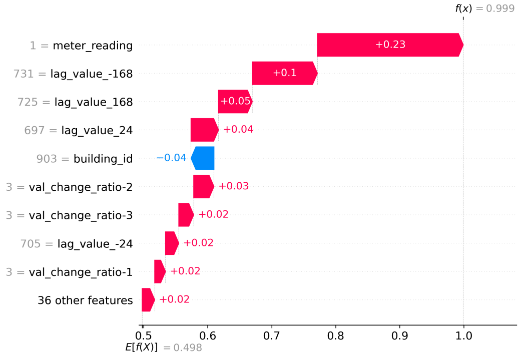

Figure 7: Waterfall plot for a specific anomaly model prediction

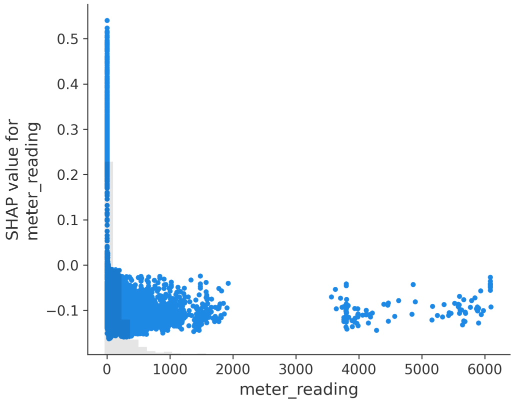

Figure 8: Scatter plot for the feature meter_reading

importance for explaining the energy prediction model. Among the features, meter reading, building_id, and lag values emerge as the most influential for the model. Figure 7 illustrates a waterfall plot for an arbitrary instance. The base prediction value is 0.49. We notice that when the meter reading value of 1, it strongly influences the prediction positively. Additionally, other crucial factors contributing to this result are the changes in reading value.

Figure 8 shows the scatter plot for the feature meter_reading. The model considers lower values of meter readings as contributing positively toward the anomaly. Figure 9 shows the scatter plot for the change in value in meter reading ($X(t) - X(t-s)$), where s is -168 hours (current - one week from now). The larger the change, the more positive impact it has on being an anomaly.

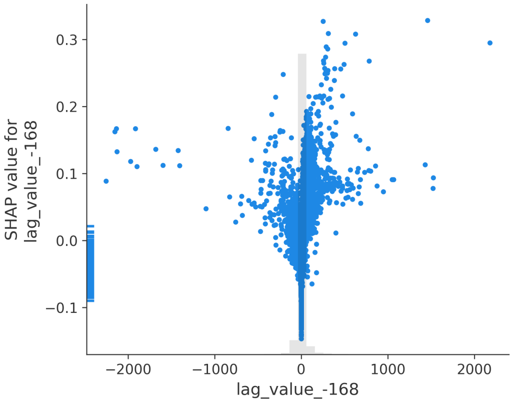

Figure 9: Scatter plot for the feature lag_value_-168

4 DISCUSSION AND CONCLUSION

Our study showcased the effectiveness of LightGBM models in predicting energy consumption and detecting anomalies within the extensive set of 200 buildings from the BDG2 dataset. Our experiments highlight the superior performance of the black-box models over baseline models for energy prediction (R-squared of 0.975 vs 0.307) and anomaly detection (AUC-ROC of 0.942 vs 0.628). However, a major hurdle in implementing LightGBM models lies in their limited interpretability. To address this, we utilized SHAP. By incorporating feature transformations and engineering directly within the model, the SHAP plots offer insights into the influence of human-interpretable features in their original scale. Key findings highlighted the positive impact of high human activity (working hours) on energy predictions and the significance of building size. For anomaly detection, meter readings and periodic fluctuations emerged as key indicators. In conclusion, the visualization capability offered by SHAP can instil confidence in building operators by providing them with a clear understanding of the model's inner workings.

REFERENCES

[1] Manoj Gulati and Pandarasamy Arjunan. 2022. LEAD1. 0: a large-scale annotated dataset for energy anomaly detection in commercial buildings. In *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*. 485–488.

[2] Srinivas Katipamula and Michael R Brambley. 2005. Methods for fault detection, diagnostics, and prognostics for building systems—a review, part I. *Hvac&R Research* 11, 1 (2005), 3–25.

[3] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[4] R Machlev, L Heistrene, M Perl, KY Levy, J Belikov, S Mannor, and Y Levron. 2022. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI* 9 (2022), 100169.

[5] Clayton Miller, Anjukan Kathirgamanathan, Bianca Picchetti, Pandarasamy Arjunan, June Young Park, Zoltan Nagy, Paul Raftery, Brodie W Hobson, Zixiao Shi, and Forrest Meggers. 2020. The building data genome project 2, energy meter data from the ASHRAE great energy predictor III competition. *Scientific data* 7, 1 (2020), 368.

[6] Saleh Seyedzadeh, Farzad Pour Rahimian, Ivan Glesk, and Marc Roper. 2018. Machine learning for estimation of building energy consumption and performance: a review. *Visualization in Engineering* 6 (2018), 1–20.