A clustering-based supervised approach for anomaly detection in building energy consumption

Beyza Nur Cansever*, and *Mehmet Zeki* Konyar

Kocaeli University, Software Engineering Department, Kocaeli, Türkiye

Abstract. Since buildings account for a significant portion of global energy consumption, energy efficiency is a critical research topic. This study proposes a novel clustering-based approach for detecting building energy anomalies. Unlike the single global model approach in the literature, the proposed method first clusters buildings according to their consumption patterns using the K-means algorithm; then, independent anomaly detection is performed for each cluster. Isolation Forest and Local Outlier Factor are used as unsupervised methods, while XGBoost, Random Forest, and LightGBM are used as supervised methods. Feature engineering was performed using building-level statistical features during the clustering phase. Experiments were conducted on the LEAD 1.0 dataset. The results show that models trained on buildings with similar consumption patterns provide significant performance improvement compared to models trained on the entire dataset. While the highest F1 score obtained by the global model was 0.572, the proposed cluster-based approach increased this value to 0.903. Among the models, XGBoost and LightGBM stood out as the best-performing models across clusters, with LightGBM achieving the highest F1 score of 0.903 in Cluster 2. Further analyses revealed that model performance improved significantly as data homogeneity increased, strengthening the effectiveness of the clustering approach.

1 Introduction

Rapid advancements in technology, particularly the Internet of Things (IoT), are transforming energy management in modern buildings, offering real-time monitoring and control. Processing and analyzing data collected from sensors makes it possible to transform buildings into smarter and more responsive systems. [1]. It is known that a significant portion of total energy consumption worldwide originates from buildings [2]. Therefore, monitoring building energy consumption with smart systems and early detection of abnormal consumption patterns has the potential to directly contribute to global energy savings.

Large volumes of time series data collected from building energy monitoring systems bring with them the problem of anomaly detection. Abnormal patterns in energy consumption can be caused by various reasons such as faulty equipment, operational irregularities, and

* Corresponding author: beyzanur.cansever@kocaeli.edu.tr

system-related malfunctions. Failure to detect anomalies in a timely manner leads to unnecessary waste of resources and negatively impacts energy efficiency [3]. Therefore, the development of anomaly detection systems is becoming increasingly essential in smart building management systems.

In the literature, studies on anomaly detection in time series data are generally addressed under supervised, unsupervised, and semi-supervised approaches. Supervised learning methods can achieve high accuracy when sufficient labelled data is available; however, they are often limited in practical applications due to challenges such as class imbalance and high labelling costs. Unsupervised and semi-supervised approaches are more widely preferred as they do not require labeled data. In this context, deep learning-based methods (LSTM, autoencoder, and GAN-based models) have gained prominence in anomaly detection due to their ability to capture complex temporal patterns in time series data. Traditional time series anomaly detection methods can have limitations in terms of accuracy and efficiency when working with complex data [4].

A significant portion of existing studies model data points with heterogeneous consumption behaviors under a single distribution, which can negatively impact anomaly detection performance in complex data structures. Accordingly, accounting for dataset heterogeneity by grouping samples with similar behavioral patterns has emerged as a noteworthy approach in recent years. Clustering-based methods offer considerable advantages in energy management and demand planning by enabling the identification of users or buildings with similar consumption behaviors [5]. Partitioning-based clustering algorithms such as K-means are among the most widely adopted methods in the literature for analyzing electricity consumption profiles, owing to their ease of implementation and high computational efficiency [6].

Clustering approaches have previously been applied to energy system data, primarily for purposes such as load profiling and customer segmentation. However, studies that directly integrate clustering into the anomaly detection process remain relatively limited in the literature. To address this gap, the proposed approach groups buildings according to their consumption patterns and performs cluster-specific anomaly detection instead of relying on a single global model.

In this study, a clustering-based framework for building energy anomaly detection is proposed. Buildings with similar consumption behaviors are grouped using the K-means algorithm, and separate anomaly detection models are trained for each cluster. For the clustering process, hourly, daily, and seasonal consumption characteristics are extracted at the building level and used as input features. Experimental results demonstrate that the proposed framework significantly improves anomaly detection performance compared to the global modeling approach. While the highest F1-score obtained using the global model was 0.572, the clustering-based approach achieved an F1-score of 0.903. These findings indicate that cluster-specific modeling is an effective strategy for handling heterogeneous building energy consumption patterns.

2 Methodology

Building energy consumption data exhibits a highly complex structure, depending on factors such as intended use, operational habits, and the physical characteristics of the building. This complexity necessitates moving away from an approach that considers all buildings under a single model and instead developing separate models for clusters formed according to consumption patterns. Evaluating buildings with different consumption patterns using the same model can lead to poor performance in both precision and recall metrics.

Based on this motivation, in our study, we propose a four-stage workflow shown in Figure 1, based on clustering according to similar consumption behaviors, on the LEAD 1.0 [7]

dataset which contains buildings with different consumption profiles. In the proposed approach, buildings are first clustered according to their consumption patterns, and then independent anomaly detection models are trained for each cluster. This section describes the four-stage workflow proposed for the study. The workflow consists of data preprocessing, feature engineering, clustering, and anomaly detection stages.

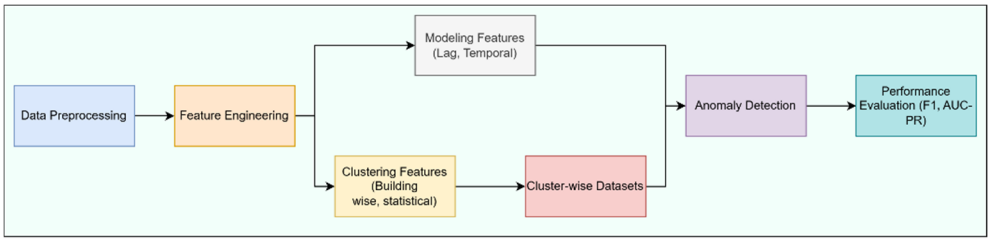


Fig. 1. Proposed workflow for cluster-based anomaly detection.

2.1 Dataset and preprocessing

This study used the Large-scale Energy Anomaly Detection (LEAD) [7] dataset which was published on Kaggle. The dataset includes data from 406 buildings, with features such as hourly measured energy consumption data, building metadata, and weather-related features spanning one year. Of the 406 buildings, 200 were allocated for training and 206 for testing. In this study, the 200-building labeled subset was used for training, validation, and testing purposes. The dataset contains approximately 8,760 hourly measurements per building, resulting in a total of approximately 1.75 million labeled records.

In the data preprocessing step, the dataset was first analyzed comprehensively. In the literature, it has been observed that buildings with missing data are excluded from the analysis [8]. Accordingly, days containing at least one missing (null) meter reading value were identified; buildings with 30 or more days of missing data were excluded from the scope of the study, as they were considered likely to adversely affect model performance. When the remaining buildings were examined, some showed missing meter readings. To maintain temporal continuity, these missing values were imputed with the building-specific median value. On the other hand, since right-skewness was observed in the meter readings, a logarithmic transformation was applied to normalize the distribution. Finally, feature selection was performed through correlation analysis, whereby features with a pairwise correlation coefficient exceeding 0.95 were removed to eliminate redundancy.

2.2 Feature engineering

Feature engineering was applied in both the clustering and anomaly detection phases. For clustering, building-level statistical and temporal features were constructed to represent the individual consumption behavior of each building. In this context, the maximum, minimum, and mean consumption values within the annual data record were calculated for each building. To better capture intra-day behavioral patterns, daily data were divided into four time periods based on the temporal classification approach proposed by Okereke et al. [5]: morning (05:00–10:00), midday (11:00–16:00), evening (17:00–22:00), and night (23:00–04:00). For each building, meter reading values were normalized individually using the building-specific maximum value. The normalized values were then used to calculate the mean consumption values at the building level. In addition to these, night-to-morning and evening-to-midday consumption ratios were calculated. Since weekday and weekend consumption values are also important for distinguishing building consumption behaviors,

weekday mean consumption, weekend mean consumption, and weekend-to-weekday consumption ratios were computed. Similarly, as seasonality plays a determining role in building consumption behavior, summer and winter mean consumption values were calculated, along with their corresponding summer-to-winter consumption ratios.

Among the features used in the clustering phase, a feature termed daily consistency was derived to measure the day-to-day consistency of energy consumption. This feature is defined as the ratio of the standard deviation to the mean (std/mean), and expresses the daily variability in a building's consumption values in a normalized manner. Higher daily consistency values indicate that consumption varies considerably from day-to-day, whereas lower values suggest a stable consumption pattern. The formula used to compute this feature is given in Equation (1):

$$\text{Daily Consistency} = \frac{\sigma_{daily}}{(\mu_{daily}+\varepsilon)} \tag{1}$$

During the feature engineering phase applied to the anomaly detection stage, lag-based features were constructed to capture the temporal dynamics of energy consumption. For each building, lag values at 1-hour, 24-hour, and 168-hour intervals were computed to reflect short-term fluctuations, daily periodicity, and weekly consumption patterns, respectively. In addition, the difference and ratio between the current meter reading and each lagged value were derived to quantify the magnitude and relative change in consumption over time.

2.3 Clustering

At this stage, the K-means algorithm [9] was chosen to group buildings with similar consumption patterns. The reasons for choosing K-means are that it produces clusters with high interpretability, works computationally efficiently in this dataset where the number of buildings is relatively limited, and consumption profiles can be easily interpreted through cluster centers.

Prior to the clustering stage, six buildings exhibiting excessive consumption patterns were identified as outliers and removed from the analysis, as their inclusion was observed to negatively affect clustering quality. The remaining building-level features were subjected to feature selection through correlation analysis; features with a correlation coefficient exceeding the threshold of 0.90 were eliminated, and standard scaling was subsequently applied. To determine the optimal number of clusters, the Silhouette score [10] and the elbow method were analyzed jointly.

For the full dataset, the Silhouette score yielded values of 0.355 for k=2 and 0.330 for k=3, which are relatively close to one another. As shown in Figure 2, the elbow method illustrates the number of clusters (k) on the x-axis and the WCSS values on the y-axis, where a distinct elbow point is observed at k=3. Based on the combined evaluation of these two analyses, k=3 was selected as the optimal cluster count.

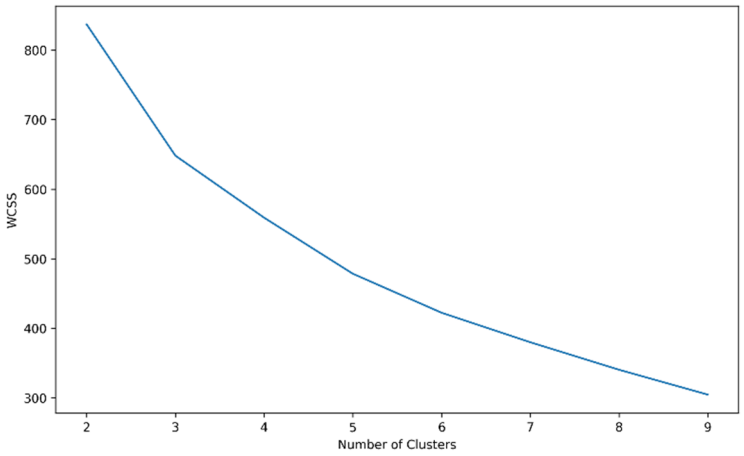


Fig. 2. Elbow plot for optimal K.

2.4 Anomaly classification

In the anomaly classification phase, separate models were trained for each cluster and their performances were compared. With this approach, instead of a single global model for the entire dataset, separate models were built within building groups with similar consumption patterns, and it was determined which model performed better in clusters with different consumption behaviors. Since class imbalance persisted after clustering, the anomaly rate in the dataset was approximately 2%, which led supervised models to lean towards the majority class. First, unsupervised methods were applied, followed by supervised approaches. Isolation Forest (IF) [11] and Local Outlier Factor (LOF) [12] were used as unsupervised methods; Random Forest [13], XGBoost [14], and LightGBM [15] were used as supervised methods. In the IF and LOF models, the contamination parameter was determined according to the anomaly rate in the training data. These models were trained only with normal data, and anomalies were attempted to be detected using the novelty detection approach. In the supervised methods, class imbalance was addressed by assigning higher weights to the minority class during training. Additionally, the decision threshold was optimized on the validation set to maximize the F1 score.

Building-based splitting was implemented to prevent data leakage during the data splitting phase. Data was split at the building level as 70% training, 20% validation, and 10% testing. This approach prevented the distribution of time series data from the same building across the training, validation, and test sets; and measured the model's generalization capacity to buildings it had never seen before.

2.5 Evaluation

Due to class imbalance in the dataset, model performance was evaluated using precision, recall, F1-score, and AUC-PR metrics rather than accuracy. Given the approximately 2% anomaly rate, high accuracy values can be misleading; a naive model that classifies all samples as normal would still achieve 98% accuracy. Therefore, AUC-PR serves as a more reliable indicator of model performance, particularly in imbalanced settings. The evaluation was conducted separately for each cluster to enable a granular analysis of model behavior across building groups with distinct consumption profiles. Two approaches are compared: a single global model trained on the entire dataset and a cluster-based modeling strategy. In

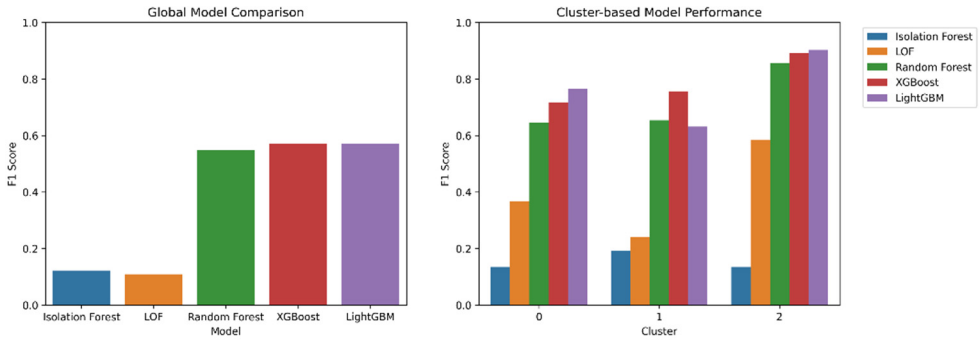
addition to these two experimental setups, a subgroup analysis focusing exclusively on Education-type buildings was performed.

3 Results

Table 1 and Figure 3 demonstrate that a single global model is limited in its ability to represent diverse consumption behaviors. In the global model, the highest F1 score and AUC-PR value were achieved by XGBoost at 0.572 and 0.539, respectively. Due to the class imbalance present in the dataset, the models were optimized using class weighting; however, despite achieving high precision, recall remained relatively low.

Table 1. Performance comparison of anomaly detection models.

Model	F1-Score	Precision	Recall	AUC-PR
Isolation Forest	0.122	0.137	0.110	0.061
LOF	0.110	0.066	0.322	0.156
Random Forest	0.548	0.967	0.383	0.510
XGBoost	0.572	0.830	0.436	0.539
LightGBM	0.571	0.840	0.433	0.538



a) b)

Fig. 3. Global (a) and cluster-based (b) model performance comparison.

As shown in Table 2 and Figure 3, the cluster-based approach incorporating the site_id feature, which also accounts for geographical context, yields a notable performance improvement over the global model. The experimental results demonstrate that model performance varied across clusters, with the optimal model differing by cluster. LightGBM achieved the highest F1 score in Cluster 0 (F1=0.766), XGBoost yielded superior performance in Cluster 1 (F1=0.756), and LightGBM outperformed other models in Cluster 2 (F1=0.903). All clusters surpassed the global model baseline (F1=0.572), indicating that the cluster-based approach provides a consistent performance improvement over the single global model. Notably, Cluster 2 exhibited the most substantial improvement, achieving an F1 score of 0.903 with LightGBM, which represents a significant gain over the global baseline. These results support the notion that geographical and climatic context is one of the primary factors governing consumption behavior.

Table 2. Best model performance per cluster.

Cluster	Best Model	F1-score	AUC-PR
Cluster 0	LightGBM	0.766	0.676
Cluster 1	XGBoost	0.756	0.703
Cluster 2	LightGBM	0.903	0.935

The results demonstrate that the cluster-based approach yields a significant performance improvement in building energy anomaly detection compared to the global model. This finding supports the premise that a single global model is insufficient to adequately capture the full diversity of consumption behaviors present in heterogeneous datasets.

When the clustering results were examined, it was observed that the grouping structure based on consumption patterns was largely shaped by geographical location and climatic conditions. A closer examination of the cluster profiles reveals distinct behavioral characteristics for each group. Cluster 0 exhibits the highest consumption levels across all time periods, with relatively low daily consistency (0.158), indicating a stable and persistently high energy demand. Cluster 1 is characterized by moderate consumption levels and the highest daily consistency value (0.386), suggesting a more variable and irregular consumption behavior, with a notably lower weekend-to-weekday ratio (0.540), implying reduced activity during weekends. Cluster 2 displays the lowest daytime consumption but the highest night-to-morning ratio (1.203), indicating that energy demand peaks during nighttime hours, which may reflect specific operational schedules or climate-driven heating and cooling loads.

To verify the effectiveness of the clustering approach performed on all buildings, an additional experiment was conducted, specifically a subgroup analysis for Education-type buildings. In this analysis, buildings in the Education category were clustered internally, separate models were trained, and the LightGBM model in the best-performing cluster within the Education-only clustering experiment achieved an F1 value of 0.988. This finding demonstrates that clustering within a homogeneous building type further improves anomaly detection performance and supports our main findings.

The lower performance of Cluster 1 compared to other clusters in the clustering performed on the entire dataset can be explained by its heterogeneous structure. When cluster-based statistics are examined, it is observed that the average daily consistency value for Cluster 1 is in a much higher range compared to other clusters. This finding shows that clustering quality directly affects anomaly detection performance and emphasizes how critical a homogeneous cluster structure is.

As observed in Figure 3, unsupervised methods such as Isolation Forest and LOF generally produced lower F1 scores compared to supervised methods. This is an expected result, as anomaly labels are manually generated on a building-by-building basis, reflecting the anomaly definition specific to each building's consumption pattern. Unsupervised methods are limited because they cannot learn this building-specific context without label information. However, the fact that unsupervised methods do not require labeled data can be considered a significant advantage in real-world applications.

In terms of the study's limitations, the building-based split approach measures the model's generalization capacity across buildings it has never seen before. While this provides a realistic assessment, determining which cluster is suitable for a new building requires a separate cluster assignment mechanism.

The impact of the clustering-based approach on performance was observed not only through the increase in metrics but also through the consistency of model behavior. Specifically, models trained on homogeneous cluster structures showed more stable decision bounds and were able to more consistently detect similar anomaly patterns across different time periods. In contrast, clusters containing buildings with diverse consumption patterns, as observed in Cluster 1, exhibited greater variability in model performance, making anomaly detection more challenging.

4 Conclusion

Real-world energy consumption data exhibit a high degree of heterogeneity due to different building types and usage habits. This makes anomaly detection difficult through a single global model and can lead to significant reductions in model performance. In particular, evaluating diverse consumption profiles under the same model is a critical problem that makes it difficult to accurately differentiate anomalous behavior. Based on this motivation, this study proposes a novel method for cluster-based anomaly classification. In this context, the traditional classification approach applied to the entire dataset is compared with the cluster-based classification approach.

According to the experimental results, the maximum model performance in the global classification was limited to an F1 score of 0.572, constrained by the heterogeneous consumption profiles present across all buildings. In contrast, the proposed cluster-based classification approach demonstrated that the best-performing cluster achieved an F1 score of 0.903. This finding confirms that evaluating buildings with similar consumption patterns within the same model significantly enhances anomaly detection performance.

From a practical standpoint, the deployment of the proposed framework requires careful consideration of how new, previously unseen buildings can be integrated. Since the clustering phase relies on building-level statistical features derived from historical consumption data, a new building must first accumulate sufficient consumption history before being assigned to an appropriate cluster via a distance-based assignment strategy. Developing an efficient mechanism for such dynamic cluster assignments represents an important direction for future work.

Considering the practical applications of the proposed method, it is predicted that grouping buildings exhibiting similar consumption patterns in existing systems to detect anomalies will both save costs and contribute positively to environmental sustainability by reducing global energy consumption. Future studies are planned to focus on more homogeneous grouping strategies, and the use of different deep learning models in anomaly detection will also be investigated. With this approach, the goal is to learn deeper patterns and achieve comprehensive anomaly detection. In addition, it is planned to identify the most distinctive features in the models by applying Explainable Artificial Intelligence (XAI) techniques.

References

1. S. Kaur, P. Chowhan, A. Sharma, A Novel Hybrid Deep Learning-Based Framework for Intelligent Anomaly Detection in Smart Meters. IEEE Access **13**, 107022 (2025). https://doi.org/10.1109/ACCESS.2025.3581257
2. Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, A. Amira, Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. Appl. Energy **287**, 116601 (2021). https://doi.org/10.1016/j.apenergy.2021.116601

3. L. Lei, B. Wu, X. Fang, L. Chen, H. Wu, W. Liu, A dynamic anomaly detection method of building energy consumption based on data mining technology. Energy **263**, 125575 (2023). https://doi.org/10.1016/j.energy.2022.125575
4. Z. Niu, K. Yu, X. Wu, LSTM-Based VAE-GAN for Time-Series Anomaly Detection. Sensors **20**, 3738 (2020). https://doi.org/10.3390/s20133738
5. G. E. Okereke, M. C. Bali, C. N. Okwueze, E. C. Ukekwe, S. C. Echezona, C. I. Ugwu, K-means clustering of electricity consumers using time-domain features from smart meter data. J. Electr. Syst. Inf. Technol. **10**, 2 (2023). https://doi.org/10.1186/s43067-023-00068-3
6. X. Liu, Y. Ding, H. Tang, F. Xiao, A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data. Energy Build. **231**, 110601 (2021). https://doi.org/10.1016/j.enbuild.2020.110601
7. M. Gulati, P. Arjunan, LEAD1.0: A Large-scale Annotated Dataset for Energy Anomaly Detection in Commercial Buildings, in Proceedings of the 13th ACM International Conference on Future Energy Systems (ACM, Virtual Event, 2022), pp. 485–488
8. A. Nukala, S. K. C. Sekhara, Y. K. Peker, M. R. Abid, A Practical Framework for Energy Fault Detection in Smart Buildings, in Proceedings of the 2025 6th International Conference on Artificial Intelligence and Robot Control (IEEE, Savannah, GA, USA, 2025), 194–201
9. S. Lloyd, Least squares quantization in PCM. IEEE Trans. Inf. Theory **28**, 129 (1982). https://doi.org/10.1109/TIT.1982.1056489
10. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53 (1987). https://doi.org/10.1016/0377-0427(87)90125-7
11. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation-based anomaly detection. ACM Trans. Knowl. Discov. Data **6**, 1-39 (2012). https://doi.org/10.1145/2133360.2133363
12. M. M. Breunig, H.-P. Kriegel, R. T. Ng, J. Sander, LOF: Identifying density-based local outliers. ACM SIGMOD Rec. **29**, 93-104 (2000). https://doi.org/10.1145/342009.335388
13. L. Breiman, Random forests. Mach. Learn. **45**, 5 (2001)
14. T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM, San Francisco, CA, USA, 2016), 785–794
15. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A highly efficient gradient boosting decision tree, in Advances in Neural Information Processing Systems (Curran Associates, Inc., 2017)