Establishment of Unsupervised Electricity Usage Behaviors Model in Energy Education

Chia-Wei Tsai[1], Cheng-Wei Hsieh[2], Li-Syuan Chen[3], Iuon-Chang Lin[2], Ying-Hsun Lai[4]*

[1] Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taiwan
[2] Department of Management Information Systems, National Chung Hsing University, Taiwan
[3] Department of Computer Science and Information Engineering, National Chung Hsing University, Taiwan
[4] Department of Computer Science and Information Engineering, National Taitung University, Taiwan
cwtsai@nutc.edu.tw, andy19990908@gmail.com, samtommycsie@gmail.com, icli@nchu.edu.tw, yhlai@nttu.edu.tw

Abstract

The detection of declining appliance energy efficiency and the analysis of common residential electricity usage patterns have garnered increasing attention in the domain of online energy education. Accurate profiling of electricity consumption behavior is instrumental in enabling precise energy management and fostering greater awareness of energy conservation. With the rapid development of artificial intelligence technology, data-driven approaches have become increasingly prevalent in identifying anomalous energy usage behaviors and forecasting energy consumption trends. This study aims to analyze real users' residential electricity consumption data using unsupervised learning techniques and explore how to detect abnormal electricity usage behavior through one-class classification algorithms. For different users, the algorithm will learn their respective normal electricity usage behavior models and identify abnormal usage behavior through these models, with the goal of improving energy efficiency and enhancing electricity safety. A particular focus is placed on high electricity-consuming appliances, such as refrigerators and air conditioners, which constitute a substantial portion of overall residential electricity usage. Through the identification of anomalous usage patterns, this approach also facilitates the monitoring and prediction of appliance health status. Consequently, the system can provide users with targeted recommendations for efficient energy utilization in everyday life. The analysis results offer users actionable insights into their electricity consumption habits, thereby supporting the development of personalized energy-saving strategies and contributing to the broader goals of energy education and sustainability.

Keywords: Unsupervised learning, Electricity usage behavior, Abnormal detecting, Energy education, Sustainability

1 Introduction

With the Sustainable Development Goals (SDGs) concept sweeping the world, there's a greater emphasis on energy conservation and reducing carbon emissions. This makes it crucial to find ways to use energy more efficiently. In households, it is vital to closely monitor and manage electricity usage in order to improve energy efficiency, minimize unnecessary electricity consumption, and enhance electrical safety in energy education. Detecting and addressing any declines in efficiency or unusual consumption patterns promptly is also of great importance. The increasing global demand for energy coupled with the effects of climate change has brought energy conservation and carbon reduction to the forefront of societal development. One key strategy for achieving these goals is improving energy efficiency in residential electricity. As urban energy consumption continues to rise, the efficiency and usage patterns of residential electricity play a crucial role in determining the energy expenses and carbon footprint of each household.

Moreover, real-time detection of electricity anomalies can significantly improve electrical safety, especially in the context of long-term care for the elderly. With the integration of modern smart home systems and IoT technologies, real-time anomaly detection in electricity usage can help identify potential issues for elderly individuals living at home, thereby improving safety and comfort. Users can understand the situation of different electricity consumption and self-consumption, which can help to build up the awareness of self-energy saving.

The emergence of smart grid technology has revolutionized the collection and analysis of residential electricity usage data. This allows for a detailed understanding of consumption patterns, identification of abnormal electricity behaviors, and prediction of future consumption trends. However, extracting valuable insights from large volumes of electricity usage data remains challenging [1]. Traditional methods for analyzing abnormal electricity usage are limited by the complexity of nonlinear usage patterns and the requirement for extensive labeled data. To address this, this study aims to use unsupervised machine learning techniques to analyze residential abnormal electricity usage data, providing real-time and valuable information for households and energy management agencies. Through the application of these

*Corresponding Author: Ying-Hsun Lai; Email: yhlai@nttu.edu.tw
DOI: https://doi.org/10.70003/160792642026112706009

techniques, the study seeks to more accurately identify abnormal electricity behaviors, ultimately contributing to enhanced safety, efficiency, and technical support for preventing and responding to abnormal events in residential electricity usage [2].

The key findings of this research can be outlined as follows, highlighting progress and real-world utility in the realm of analyzing residential electricity data.

- Using real residential electricity datasets: By incorporating real-world electricity datasets, we can ensure that our approach accurately mirrors the intricate and diverse patterns of electricity usage observed in everyday life. This allows us to closely replicate actual electricity consumption behaviors, mitigating the risks of errors and biases that are often present when training models in synthetic environments. Utilizing real-world electricity data not only bolsters the credibility of the model but also elevates its ability to adapt and perform effectively in real-world application scenarios [3].
- Using unsupervised learning to train the model: In supervised learning, machine learning models are typically trained using a large amount of labeled data. However, acquiring and labeling such data can be costly and, at times, impractical. This study utilizes unsupervised learning techniques to train the model without relying on pre-labeled data. By doing so, the model is able to understand typical patterns of electricity usage and detect abnormal behaviors without the need for labeled data. This method broadens the model's adaptability and its ability to identify anomalies in unfamiliar scenarios, particularly excelling at recognizing new and unusual electricity consumption patterns. Not only does this approach enhance the efficiency and precision of anomaly detection, but it also establishes a dependable analytical foundation for the safety and management of electricity usage [4].
- Improving safety and energy-saving of residential electricity: Through the extensive analysis of residential electricity data, this study offers valuable insights aimed at improving electricity safety and optimizing electricity management within households. By closely examining abnormal usage patterns and potential performance degradation, this research facilitates real-time monitoring and identifies opportunities for conserving energy in residential electricity consumption. Not only does this approach aid in the prevention of electricity-related accidents and raise the safety standards of household electricity usage, but it also delivers practical energy-saving recommendations for residents, thereby fostering efficient energy utilization. Ultimately, the impact of this achievement is substantial in creating a safer and more energy-efficient environment within residential electricity systems [5].

2 Related Work

2.1 Anomaly Detection

In the realm of data analysis, anomalies are identified as individual or groups of data points within a dataset that significantly diverge from the typical range of most data or substantially surpass the anticipated behavioral patterns for that dataset. Consequently, these anomalies conspicuously manifest or draw attention in data visualization or analysis outcomes. Such anomalous points are frequently deemed as pivotal indicators, possibly signifying pertinent information or cautionary indications, thereby inferring potential system malfunctions or other exigencies necessitating attention. Thus, they assume a pivotal role in data surveillance and decision-making procedures.

Anomaly detection denotes the task of discerning atypical data points in a dataset that do not adhere to normal behavioral patterns. Time series data encompasses a succession of data points arranged chronologically, commonly encountered in domains such as finance, healthcare, and engineering. The essence of anomaly detection lies in pinpointing those data points that deviate from the envisaged patterns. These anomalies may arise from system dysfunctions or other irregular conditions [6]. The extensive research in this domain encompasses diverse methodologies and techniques. By identifying and scrutinizing these deviations, we can attain profound insights into the underlying reasons they may be concealed, thereby enabling us not only to promptly detect and rectify issues but also to preclude potential adversities and even unearth valuable new insights embedded within the data.

2.2 Methods of Anomaly Detection

Anomaly detection involves identifying data points that deviate from the majority of the data in a dataset [6]. Traditionally, statistical-based and distance-based methods have been used for this purpose. However, with the rapid advancement of machine learning, more sophisticated techniques have been developed, allowing for more flexible and precise anomaly detection. In the realm of machine learning, there are several commonly used methods for anomaly detection, including Supervised anomaly detection, Semi-supervised anomaly detection, and Unsupervised anomaly detection [7].

2.2.1 Supervised Anomaly Detection

In the realm of supervised anomaly detection methods, the typical approach involves training two distinct models using a dataset that contains both normal and abnormal labels. One of these models is specifically trained to identify normal data, while the other is developed to recognize abnormal data. It is important to note that there are several challenges associated with this approach. Data imbalance is a significant concern, as the number of abnormal data samples is often much lower than that of normal data samples. This disparity can potentially lead the model to exhibit bias towards the more frequently occurring class. Another critical issue lies in the quality of labeling, particularly in the context of abnormal data. Accurately determining abnormal data labeling is often

more challenging, which can consequently impact the effectiveness of model training.

While supervised learning remains a potent method in anomaly detection, the quality and quantity of training data play pivotal roles in determining the efficacy of the resulting models. In practical applications, it is essential to explore and implement robust data balancing strategies, such as the Synthetic Minority Over-sampling Technique (SMOTE) [8], to address the scarcity of abnormal samples. Additionally, improving the accuracy of labeling becomes crucial, and may necessitate a more meticulous labeling process or expert intervention to ensure the quality of data labeling. These enhancements are instrumental in bolstering the model's capability to identify abnormal conditions, thereby facilitating more effective anomaly detection [9-10].

2.2.2 Semi-supervised Anomaly Detection

In semi-supervised anomaly detection, the training process typically involves using labelled in-stances of the normal class, while the remaining unlabelled data points are considered as potential anomalies. This method offers the significant advantage of not requiring labelled data for abnormal instances, making it more broadly applicable in real-world scenarios.

Semi-supervised learning is particularly well-suited for situations where abnormal data is scarce or challenging to precisely define, as it necessitates only a small amount of labelling. However, a key challenge of this approach is its dependence on the representativeness of the normal samples. If the selected normal samples fail to accurately capture all normal behaviors, it can result in a higher false positive rate. Therefore, ensuring the utilization of a sufficiently diverse and comprehensive set of normal samples is critical for enhancing the accuracy of this method [11-12].

2.2.3 Unsupervised Anomaly Detection

In unsupervised anomaly detection, the model does not require pre-labelled training data for training, which can save a significant amount of time and resources. This is particularly advantageous in cases where acquiring labelled data is costly or technically infeasible. These methods operate on the premise that, in most cases, the frequency of normal data occurrences is much higher than that of anomalous data. However, it's important to note that if this assumption is inaccurate, using this approach may lead to a high false positive rate.

Unsupervised anomaly detection offers greater flexibility, especially in handling unknown or newly emerging anomaly types. It is also highly useful in situations where obtaining a large amount of labelled data is not feasible, allowing for rapid deployment in various environments. This approach can be particularly valuable in scenarios where the nature of anomalies is constantly evolving or where extensive labelled data is not readily available [13-14].

3 Research Methods

The growing focus on energy efficiency highlights the critical importance of promptly identifying abnormal usage patterns to ensure energy security and optimize energy management. This research introduces a methodology designed to offer a comprehensive approach for analyzing home energy management by detecting anomalies on an hourly basis. It is important to emphasize that the dataset utilized in this study comprises real user electricity consumption data. This inclusion not only enhances the practical relevance of the study but also bolsters the credibility and reliability of the model. Leveraging real data in this manner ensures that the research findings are firmly grounded in empirical evidence, making them more applicable in real-world scenarios. Ultimately, this approach equips users with practical, effective, and data-supported insights [15].

3.1 Feature Engineering

To achieve hour-by-hour anomaly detection, we employ a segmentation method to divide the time series power consumption data, which is recorded every minute, into different segments based on a 60-minute time window. This approach allows for the analysis of both current and historical data, enabling moment-by-moment anomaly detection without relying on future data. Figure 1 is a schematic diagram of the 60-minute time window.

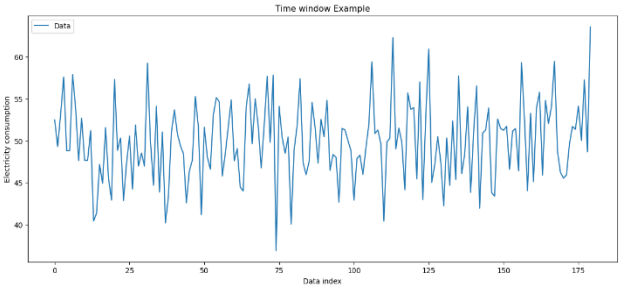

Figure 1. 60-Minute time window schematic

After segmenting the time windows, we extract various features related to electricity consumption behavior. These features include the average electricity consumption, standard deviation of electricity consumption, maximum electricity consumption, and minimum power usage. Each of these characteristics provides valuable insights into the patterns of power consumption. For instance, the average power usage reveals the typical power consumption level within the window, while the standard deviation reflects the variability of power usage, indicating fluctuations and instability during that period. Additionally, analyzing maximum and minimum usage offers insights into the extremes within the window, helping in the identification of peaks and valleys in power usage.

In Table 1, we provide a comprehensive description of the characteristics of the time window, detailing how each feature contributes to our understanding of power consumption behavior.

Table 1. Description of time window feature values

Feature value	Description
w_i	The i-th data entry of electricity consumption within the specified time window, with a total of 60 entries.
Average usage	Mean electricity consumption during the specified period
Standard deviation of usage	Reflects the variability and changes in electricity usage during the time window
Maximum usage	Maximum electricity consumption during the specified period
Minimum usage	Minimum electricity consumption during the specified period

3.2 Clustering of Electricity Usage Patterns

Upon processing the feature values, we employ the K-means algorithm to cluster electricity usage behaviors, grouping similar behaviors together [16-17]. The number of clusters is determined using the Silhouette Coefficient [17], which measures the quality of clustering by considering both in-tra-cluster similarity and inter-cluster differences. This coefficient ranges from -1 to 1, with values closer to 1 indicating better clustering and those closer to -1 indicating poorer clustering.

We have calculated silhouette coefficients for varying numbers of clusters using the electricity usage data of five users. The results, illustrated in Figure 2, show that the highest silhouette coefficient for each user's electricity usage behavior occurs when the data is divided into two clusters. Furthermore, the coefficient tends to decrease as the number of clusters increases. Based on this analysis, the electricity usage data for each user will be divided into two clusters for experimentation, ensuring reasonable and interpretable clustering results.

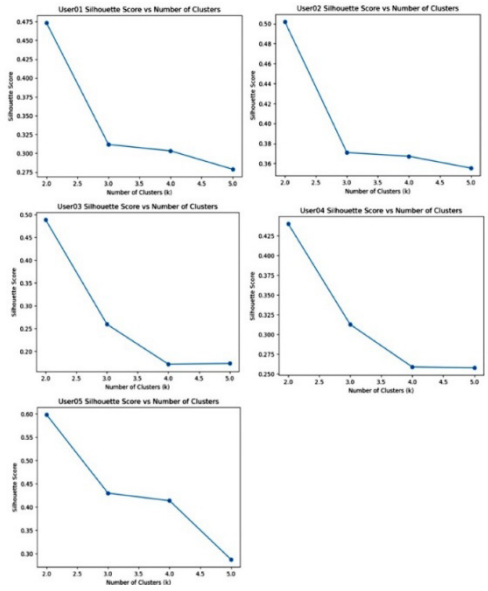

Figure 2. Number of clusters and Silhouette coefficients relationship for 5 users

3.3 Abnormal Electricity Usage Behaviors

To test the effectiveness of our model in identifying abnormal power usage patterns, we con-ducted hour-by-hour anomaly detection simulations. The simulations cover five different types of abnormal behavior: sudden spikes, sudden drops, sustained high usage, abnormal standby mode, and dramatic changes in power usage [18]. These scenarios are designed to replicate potentially anomalous behavior such as unexpected surges and dips in power usage, prolonged periods of high usage, ir-regular standby modes, and large fluctuations in power usage. By simulating these anomalous be-haviors, we are able to thoroughly evaluate the model's ability to detect abnormal power patterns, ensuring that it remains effective at identifying anomalies under a variety of conditions. For detailed information on the five categories of abnormal behavior, please refer to Table 2 below.

Table 2. Description of abnormal behaviors

Abnormal behavior type	Description
Sudden spike	An abrupt surge in electricity consumption within a specific timeframe, potentially stemming from equipment leakage or malfunction.
Sudden drop	A sudden and substantial reduction in the consumption of electricity over a specific period, potentially resulting from the inadvertent shutdown of equipment.
High usage	Unusually high electricity consumption continues for an extended period, possibly caused by equipment overload or an ongoing malfunction.
Abnormal standby	It appears that almost no electricity is being used during this time period. This could mean that the equipment is not being used when expected, or there may be other reasons for the electricity interruption.
Dramatic variability	The electricity usage pattern shows rapid and frequent fluctuations over a short period, suggesting highly unstable consumption trends. This may be due to grid malfunctions or erratic start/stop cycles of equipment.

In our study, we conducted a series of tests to evaluate how well the model could detect different levels of abnormal changes in data. We deliberately created abnormal variations in data at various magnitudes, ranging from minor fluctuations to extreme alterations, to analyze the model's accuracy. These experiments were designed to assess the model's resilience when encountering anomalies of varying intensities and to confirm its ability to effectively recognize and respond to abnormal situations in real-world scenarios. Table 3 provides a breakdown of the abnormal changes in electricity usage behavior across

different periods, illustrating how electricity consumption is modified by varying degrees during specific timeframes to replicate diverse abnormal electricity scenarios.

3.4 Abnormal Electricity Usage Behaviors

In our approach to training a model that detects abnormal power consumption on an hourly basis, we choose Isolation Forest as our primary anomaly detection model [19]. Isolation Forest is an un-supervised model that is well-suited for identifying anomalies and excels at processing large amounts of data quickly. Only normal data is required in the training phase, and the model can effectively isolate abnormal data during testing.

Based on the results of electricity consumption behavior clustering, we divided each user's behavior into two different clusters. We then developed separate detection models for each user cluster (Figure 2) to ensure we have a comprehensive understanding of normal power usage behavior within each cluster. This tailored approach improves the sensitivity and accuracy of anomaly detection. By training independent models for each cluster, we are able to analyze different power usage behavior patterns. This helps to look more closely at specific behavioral characteristics, such as normal daily electricity usage compared to peak usage periods, which typically show higher consumption levels but still within normal limits.

To further improve our detection performance, we leverage Optuna to search for the best parameters for each user model. Optuna is an open source hyperparameter optimization framework designed to automate the process of fine-tuning machine learning models [20]. This allows each in-dividual model to better adapt to the unique characteristics of its corresponding user population, thereby reducing false positives while maintaining high sensitivity.

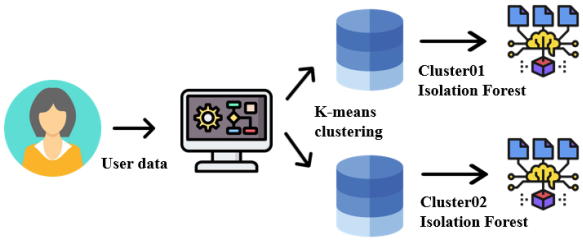

Figure 3. Individual training models for each cluster of each user

4 Research and Experimental Results

4.1 Dataset and Experimental Platform

The data used for this study consists of real-time electricity consumption data from households in Taiwan, collected through smart meters. These data provide a detailed record of total electricity usage for each residence and are collected at 1-minute intervals, with the electricity consumption measured in watts (W). The data spans from June 1, 2023, to February 29, 2024, covering a period of 9 months and totaling approximately 394,560 data points. This high-frequency and long-term data collection allows for a comprehensive analysis of daily power usage patterns and the identification of abnormal electricity consumption behaviors. The study will involve experiments conducted on the electricity consumption data of 5 different users, with individual independent models trained for each user to accurately capture their unique electricity consumption habits and behaviors.

4.2 Confusion and Performance Metrics

The Confusion Matrix plays a crucial role in machine learning as it helps in evaluating the performance of classification models. It is a visual representation of the model's predictions compared to the actual outcomes, often used in scenarios where there are two classes to predict. In the provided figure (Figure 3), the confusion matrix visually breaks down the model's predictions into four main categories:

(1) True Positive (TP): The model correctly identifies a positive case. This means the model predicts the class as positive, and the data is indeed positive.
(2) False Positive (FP): The model incorrectly identifies a positive case. This occurs when the model predicts the class as positive, but the data is actually negative. This error is sometimes referred to as a "false alarm."
(3) True Negative (TN): The model correctly identifies a negative case. This indicates that the model predicts the class as negative, and the data is indeed negative.
(4) False Negative (FN): The model incorrectly identifies a negative case. This happens when the model predicts the class as negative, but the data is actually positive. This error is sometimes referred to as a "miss."

Understanding and evaluating the predictive performance of a model relies on four fundamental elements. For instance, the model's accuracy is increased when there are more true positives and true negatives. Conversely, minimizing the number of false positives and false negatives is essential to reduce prediction errors. he confusion matrix, which was previously explained, has other significant purposes such as the calculation of statistical metrics like Precision, Recall, and the F1 Score. Here's an introduction to these three important metrics.

Precision is a key measure used to evaluate the effectiveness of a classification model in identifying positive classes. It indicates the ratio of true positives to the total number of instances that the model has classified as positive. In the context of anomaly detection, precision measures the accuracy of identifying anomalies, revealing the proportion of correctly identified anomalies out of all cases flagged as anomalies by the model. The precision of finding anomalies can be calculated using the following formula.

$$Precision=\frac{True\ Positives(TP)}{True\ Positives(TP)+False\ Positives(FP)} \quad (1)$$

Recall is a metric used to measure a classification model's ability to correctly identify all actual positive cases. It focuses on how well the model captures positive labels, which is the proportion of true positive cases that the model can recognize. In the context of anomaly detection, recall indicates the model's success in identifying real anomalies among all true anomaly events. The formula for calculating recall is:

$$Recall=\frac{True\ Positives(TP)}{True\ Positives(TP)+False\ Negatives(FN)} \quad (2)$$

The F1-score is a key evaluation metric used to gauge the overall effectiveness of a classification model. It takes into account both Precision and Recall, offering a single numerical value that encapsulates the model's performance. In the realm of anomaly detection, the F1-score strikes a balance between accurately identifying anomalies and minimizing the prevalence of missed anomalies. This means that a high F1-score model excels not only in precision but also in accurately recalling true anomalies. The F1-score is calculated using the following formula:

$$F1-score=\frac{2\times Precision\times Recall}{Precision+Recall} \quad (3)$$

4.3 Experimental Results

In our anomaly detection system, we analyze electricity usage data recorded every minute and create 60-minute segments of behavior. We use the K-means algorithm to cluster the electricity consumption behavior of each use. Each user's behavior is divided into two groups, and a separate anomaly detection model is trained for each group. We also tested various abnormal power usage scenarios, and the evaluation results of five users are shown in Table 4. The F1-score of each user's model is consistently over 0.90 under abnormal multiple combinations 2 and 3, indicating that each model performs quite well for each user. For the abnormal multiple combination 1, the F1-scores of most models are over 0.90, except for User01-cluster2 and User02-cluster2. This is because the electricity usage behavior of these two users exhibits significant randomness, making it difficult for the model to identify anomalies with lower degrees of abnormality.

Table 3. Combinations of abnormal multipliers

	Abnormal multiple combination 1	Abnormal multiple combination 2	Abnormal multiple combination 3
Sudden spike	Multiply electricity usage by 5 times for w_{21}~w_{40}	Multiply electricity usage by 7 times for w_{21}~w_{40}	Multiply electricity usage by 10 times for w_{21}~w_{40}
Sudden drop	Multiply electricity usage by 5 times for w_1~w_{20}, w_{41}~w_{60}, multiply electricity usage by 0.5 times for w_{21}~w_{40}	Multiply electricity usage by 7 times for w_1~w_{20}, w_{41}~w_{60}, multiply electricity usage by 0.7 times for w_{21}~w_{40}	Multiply electricity usage by 7 times for w_1~w_{20}, w_{41}~w_{60}, electricity usage remains unchanged for w_{21}~w_{40}
Consistently high usage	Multiply electricity usage by 5 times for w_1~w_{60}	Multiply electricity usage by 7 times for w_1~w_{60}	Multiply electricity usage by 10 times for w_1~w_{60}
Abnormal idle	Set electricity usage to random values between 1-10 for w_1~w_{60}	Set electricity usage to random values between 1-10 for w_1~w_{60}	Set electricity usage to random values between 1-10 for w_1~w_{60}
Anomaly volatile	Randomly multiply electricity usage by 5 or 0.2 for w_1~w_{60}	Randomly multiply electricity usage by 7 or 0.15 for w_1~w_{60}	Randomly multiply electricity usage by 10 or 0.1 for w_1~w_{60}

Table 4. Evaluation results for different user clusters and scenarios

User clusters	Scenarios	F1-score	Precision	Recall
User01-cluster 1	Abnormal multiple combination 1	0.96	0.95	0.97
	Abnormal multiple combination 2	0.97	0.95	1.00
	Abnormal multiple combination 3	0.98	0.96	1.00
User01-cluster 2	Abnormal multiple combination 1	0.87	0.83	0.92
	Abnormal multiple combination 2	0.91	0.84	0.99
	Abnormal multiple combination 3	0.91	0.84	0.99
User02-cluster 1	Abnormal multiple combination 1	0.92	0.87	0.98
	Abnormal multiple combination 2	0.93	0.87	1.00
	Abnormal multiple combination 3	0.95	0.91	1.00
User02-cluster 2	Abnormal multiple combination 1	0.74	0.77	0.71
	Abnormal multiple combination 2	0.90	0.90	0.88
	Abnormal multiple combination 3	0.91	0.91	0.92
User03-cluster 1	Abnormal multiple combination 1	0.96	0.92	0.99
	Abnormal multiple combination 2	0.98	0.97	1.00
	Abnormal multiple combination 3	0.98	0.97	1.00
User03-cluster 2`	Abnormal multiple combination 1	0.91	0.84	1.00
	Abnormal multiple combination 2	0.91	0.84	1.00
	Abnormal multiple combination 3	0.91	0.84	1.00
User04-cluster 1	Abnormal multiple combination 1	0.93	0.91	0.96
	Abnormal multiple combination 2	0.97	0.95	0.99
	Abnormal multiple combination 3	0.98	0.95	1.00
User04-cluster 2	Abnormal multiple combination 1	0.89	0.83	0.95
	Abnormal multiple combination 2	0.91	0.84	1.00
	Abnormal multiple combination 3	0.91	0.84	1.00
User05-cluster 1	Abnormal multiple combination 1	0.76	0.83	0.71
	Abnormal multiple combination 2	0.90	0.89	0.89
	Abnormal multiple combination 3	0.91	0.93	0.90
User05-cluster 2	Abnormal multiple combination 1	0.91	0.84	0.99
	Abnormal multiple combination 2	0.91	0.84	1.00
	Abnormal multiple combination 3	0.91	0.84	1.00

5 Conclusions

The escalating global energy demand, coupled with the urgent need for energy conservation and carbon emission reduction, has placed energy management and optimization at the forefront of scientific inquiry. In the domain of residential electricity usage, accurately monitoring consumption patterns and promptly identifying anomalous behaviors are critical to enhancing energy efficiency and ensuring electrical safety—key objectives in the context of energy education and awareness.

By leveraging actual residential electricity consumption data, the study utilizes isolation forests to identify abnormal electricity usage behaviors. The findings indicate that this approach effectively recognizes irregular consumption patterns, thereby offering valuable insights for energy management. Furthermore, the study introduces an hourly anomaly detection method that enables real-time monitoring of electricity usage, thereby enhancing safety and management efficiency.

The key contributions and outcomes of this study include the utilization of real electricity consumption data to enhance the model's credibility and practicality. By avoiding errors and biases inherent in simulated environments, this approach provides a more accurate reflection of daily electricity usage patterns and behaviors, thereby boosting the model's adaptability and effectiveness in real-world applications. Additionally, the study demonstrates the efficacy of unsupervised learning methods in accurately identifying abnormal consumption patterns without requiring a large amount of labeled data, making it highly suitable for practical applications. Lastly, the proposed hourly anomaly detection method, which utilizes K-means clustering, enables real-time monitoring of abnormal usage behaviors, thereby significantly improving electrical safety.

The research utilizes advanced unsupervised machine learning techniques to effectively identify anomalies in residential electricity usage.

Despite the promising results, there are various areas that require further investigation and enhancement for future research. Future research opportunities include exploring additional unsupervised learning methods beyond Isolation Forest and K-means clustering. For instance, one-class SVM [21] and Autoencoders [22] offer stronger feature extraction and pattern recognition capabilities that can address more intricate electricity usage behavior patterns. Furthermore, there is potential for further investigation into analyzing and classifying the causes of abnormal electricity usage behavior. This includes distinguishing various types of abnormal behavior, such as faults, changes in electricity usage habits, and illegal electricity usage. By taking this approach, users can receive more precise recommendations and measures, thereby enhancing the self-awareness of electricity saving education.

Acknowledgements

Our research team gratefully acknowledges the financial support from Energy Administration, Ministry of Economic Affairs, R.O.C., for this study (Grant No. 113-E0208).

References

[1] S. V. Oprea, A. Bâra, F. C. Puican, I. C. Radu, Anomaly detection with machine learning algorithms and big data in electricity consumption, *Sustainability*, Vol. 13, No. 19, Article No. 10963, October, 2021.

[2] X. Wang, Z. Yao, M. Papaefthymiou, A real-time electrical load forecasting and unsupervised anomaly detection framework, *Applied Energy*, Vol. 330, Part A, Article No. 120279, January, 2023.

[3] J. Kelly, W. Knottenbelt, The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes, *Scientific Data*, Vol. 2, pp. 1-14, March, 2015.

[4] A. Aligholian, M. Farajollahi, H. Mohsenian-Rad, Unsupervised learning for online abnormality detection in smart meter data, *2019 IEEE Power & Energy Society General Meeting (PESGM)*, Atlanta, GA, USA, 2019, pp. 1-5.

[5] K. D. Trung, C. V. Tuyet, Elasticities of Residential Electricity Demand in Vietnam, *2021 International Conference on Smart City and Green Energy (ICSCGE)*, Hangzhou, China, 2021, pp. 103-107.

[6] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, *ACM Computing Surveys (CSUR)*, Vol. 41, No. 3, pp. 1-58, July 2009.

[7] A. B. Nassif, M. A. Talib, Q. Nasir, F. M. Dakalbab, Machine learning for anomaly detection: A systematic review, *IEEE Access*, Vol. 9, pp. 78658-78700, May, 2021.

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321-357, June, 2002.

[9] Y. F. Tan, W. N. Tan, R. El-Hadad, A. Buhari, Predicting Power Consumption Anomaly Using Statistical and Supervised Machine Learning Techniques, *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, Chongqing, China, 2022, pp. 671-676.

[10] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, A. Gómez-Expósito, Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning, *IEEE Transactions on Smart Grid*, Vol. 10, No. 3, pp. 2661-2670, May 2019.

[11] Y. Zhou, W. Hu, Y. Min, L. Zheng, B. Liu, R. Yu, Y. Dong, A semi-supervised anomaly detection method for wind farm power data preprocessing, *2017 IEEE Power & Energy Society General Meeting*, Chicago, IL, USA, 2017, pp. 1-5.

[12] X. Wang, I. Yang, S. H. Ahn, Sample Efficient Home Power Anomaly Detection in Real Time Using Semi-Supervised Learning, *IEEE Access*, Vol. 7, pp. 139712-139725, September 2019.

[13] M. Z. H. Jesmeen, G. R. Murthy, J. Hossen, J. Ganesan, A. Abd Aziz, K. Tawsif, Detecting abnormal electricity usage using unsupervised learning model in unlabeled data, *International Journal of Advanced and Applied Sciences*,

Vol. 8, No. 9, pp. 102-111, September, 2021.

[14] S. M. Kaddour, M. Lehsaini, Electricity consumption data analysis using various outlier detection methods, *International Journal of Software Science and Computational Intelligence (IJSSCI)*, Vol. 13, No. 3, pp. 12-27, July-September, 2021.

[15] J. S. Chou, A. S. Telaga, Real-time detection of anomalous power consumption, *Renewable and Sustainable Energy Reviews*, Vol. 33, pp. 400-411, May, 2014.

[16] M. Ahmed, R. Seraj, S. M. S. Islam, The k-means Algorithm: A Comprehensive Survey and Performance Evaluation, *Electronics*, Vol. 9, No. 8, Article No. 1295, August, 2020.

[17] Y. Amril, A. L. Fadhilah, Fatmawati, N. Setiani, S. Rani, Analysis Clustering of Electricity Usage Profile Using K-Means Algorithm, *IOP Conference Series: Materials Science and Engineering*, Vol. 105, pp. 1-8, 2016.

[18] W. Zhang, X. Dong, H. Li, J. Xu, D. Wang, Unsupervised Detection of Abnormal Electricity Consumption Behavior Based on Feature Engineering, *IEEE Access*, Vol. 8, pp. 55483-55500, March, 2020.

[19] F. T. Liu, K. M. Ting, Z. H. Zhou, Isolation Forest, *2008 Eighth IEEE International Conference on Data Mining*, Pisa, Italy, 2008, pp. 413-422.

[20] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, *25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA, 2019, pp. 2623-2631.

[21] K. L. Li, H. K. Huang, S. F. Tian, W. Xu, Improving one-class SVM for anomaly detection, *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, Xi'an, China, 2003, pp. 3077-3081.

[22] C. Zhou, C. R. Paffenroth, Anomaly detection with robust deep autoencoders, *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax, Canada, 2017, pp. 665-674.

Biographies



Chia-Wei Tsai received the Ph.D. degree in Computer Science and Information Engineering from National Cheng Kung University, Tainan, Taiwan, in 2011. He is currently an Associate Professor of National Taichung University of Science and Technology. His research interests include quantum cryptography, quantum information & computation, and anomaly detection.



Cheng-Wei Hsieh graduated from the master's program at National Chung Hsing University. His research centers on machine learning, deep learning, and anomaly detection, with a strong focus on applying these approaches to anomalous behavior analysis and prediction.



Li-Syuan Chen is currently a Ph.D. student in the Department of Computer Science and Engineering at National Chung Hsing University. His expertise includes artificial intelligence, image processing, and big data analysis.



Iuon-Chang Lin received the Ph.D. in Computer Science and Information Engineering in March 2004 from National Chung Cheng University, Chiayi, Taiwan. He is currently a professor and chair of the Department of Management Information Systems, National Chung Hsing University, Taichung, Taiwan. His current research interests include electronic commerce, information security, blockchain security, and cloud computing.



Ying-Hsun Lai received the Ph.D. degree in computer and information sciences from National Cheng Kung University, Tainan, Taiwan. He is currently a Professor with the Department of Computer Science and Information Engineering, National Taitung University, Taitung. His research interests include engineering education, artificial intelligence and industrial Internet-of-Things security technology and applications, STEAM education, and multicultural education.