COMPARISON OF MACHINE LEARNING CLUSTERING ALGORITHMS FOR ANALYSING ELECTRICITY USAGE PATTERNS IN CAMPUS AREAS

Diya Namira Purba[1)*], Muhammad Ridha[2)], Rida Indah Fariani[3)], Harkiapri Yanto[4)]

[1,2]Informatics Management, [3]Software Engineering, [4]Mechanical Engineering

[1,2,3,4]Astra Polytechnic

E-mail: diya.namira@polytechnic.astra.ac.id[1)], m.ridha@polytechnic.astra.ac.id[2)], Rida.Fariani@polytechnic.astra.ac.id[3)], harkiapri.yanto@polytechnic.astra.ac.id[4)]

Abstract

Electricity consumption in campus environments varies based on building functions, occupancy patterns, and time-of-day usage. Understanding these variations is essential for efficient energy management. Uncontrolled electricity use often results in high operational costs, highlighting the need for accurate methods to uncover consumption patterns. This study analyzes electricity consumption data from multiple campus buildings at a polytechnic in Jakarta during 2023 and 2024. Each dataset consists of six columns and 365 rows in a year. Since the data is unlabeled, three clustering algorithms: K-Means, Hierarchical Clustering, and DBSCAN are applied to identify usage patterns across campus areas. Pre-processing included imputation and normalization, followed by clustering. Cluster quality was evaluated using the Silhouette Score. A key novelty of this study is the year-to-year comparative analysis, showing that clustering performance can vary significantly depending on data structure and noise. The 2023 dataset (dataset 1) achieved the highest Silhouette Score of 0.48 using DBSCAN, while the 2024 dataset (dataset 2) produced the best result with Hierarchical Clustering at 0.53. These results emphasize the importance of selecting clustering methods based on data characteristics and temporal context. The findings contribute to developing adaptive, data-driven strategies for managing energy use in non-residential settings, particularly in educational institutions like campuses.

Keywords- *electricity consumption, campus area, clustering algorithm, machine learning*

1. INTRODUCTION

Understanding electricity usage patterns is crucial for optimizing energy consumption and implementing effective energy management strategies in non-residential settings, such as campus area. Identifying distinct usage patterns can lead to better energy-saving initiatives and inform policy decisions aimed at reducing overall consumption in these areas [1]. Recognizing these patterns not only enhances operational efficiency but also supports sustainability efforts by minimizing energy waste and promoting responsible consumption practices [1].

The increasing emphasis on energy efficiency and sustainability has led researchers and practitioners to explore innovative solutions for managing electricity consumption across campus areas. As campuses rely on a diverse array of electrical appliances and technologies, it is essential to analyze overall electricity usage data to identify patterns and trends. The application of machine learning techniques, particularly clustering algorithms, provides a valuable approach to examining this aggregated data. Clustering groups similar consumption profiles, allowing for the identification of distinct usage behaviors across the campus. This comprehensive understanding facilitates targeted interventions and informed decision-making regarding energy distribution and conservation strategies for the entire campus environment [1], [2], [3].

Beyond analysis, clustering algorithms also play a pivotal role in predictive modeling to enhance energy management strategies. Techniques such as agglomerative hierarchical clustering allow stakeholders to not only categorize current usage patterns but also forecast future trends based on historical data. For instance, a study

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96

analyzed electricity consumption across various campus buildings, employing K-Means clustering to group buildings based on daily power usage [4]. The study developed a combined forecasting model that outperformed traditional methods like LSTM and SVR, enabling more accurate short-term predictions of electricity consumption. This approach facilitated tailored energy-saving strategies and informed demand-side management initiatives. As organizations increasingly adopt machine learning techniques for energy management, the potential for enhanced analysis of electricity usage patterns becomes evident. By utilizing advanced clustering algorithms, businesses can effectively categorize and analyze historical energy consumption data, allowing for the identification of distinct usage patterns across various non-residential sectors. For example, K-means clustering and agglomerative hierarchical clustering can reveal nuanced consumption behaviors that inform tailored energy management strategies [5].

Several related studies have implemented clustering algorithms to analyze electricity consumption. In 2022, a study on electricity pattern analysis by clustering domestic load profiles using the discrete wavelet transform. This study utilized the Manhattan dataset, implementing agglomerative hierarchical clustering. It also implemented cluster validity indices (CV). The result was evaluated using three evaluation methods, including the silhouette coefficient and the Calinski-Harabasz index. The result indicated that the clustering algorithms, when combined with discrete wavelet transform, improved clustering performance [6]. Another related study is about the K-means clustering of electricity consumers using time-domain features from smart meters. The data used in this study were collected from energy consumption data of 5,667 London households that participated in the UK power networks. The algorithm used the K-means clustering algorithm. The result of the algorithm helps the power supplier identify consumption behaviors based on the extracted temporal features [7].

A study on electricity consumption patterns using SOM-based two-level clustering of residential households, proposed by Chavda et al. This study utilized electric power consumption measurements from 6,445 households. This study implemented several clustering algorithms, including K-means, Gaussian Mixture clustering, MiniBatch, KMeans, Agglomerative clustering, and Spectral clustering, on the D-EC data of electrical usage. The study identified four distinct clusters of electricity consumption patterns among residential households, categorizing consumers into Low Usage Consumers, Moderate Usage Consumers, Elevated Usage Consumers, and Extravagant Usage Consumers based on their daily consumption ranges [8]. Another related study on the clustering analysis of electricity behavior was proposed by Liu et al. The dataset used in this study was collected from industry electricity usage measurements. This study employed three algorithms, including K-means, fuzzy, and neural network clustering. Lastly, this study classifies them based on their electricity consumption behavior by analyzing load curves [9].

From all previous studies, the application of clustering algorithms in analyzing electricity usage patterns is pivotal for deriving meaningful insights that can inform energy management strategies. The right clustering technique can significantly enhance the analysis, enabling stakeholders to identify and understand distinct consumption behaviors effectively. This paper aims to compare three prominent clustering algorithms: K-means, hierarchical clustering, and DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Each of these methods offers unique advantages and may yield varying results depending on the nature of the electricity usage data.

K-means clustering is renowned for its efficiency in handling large datasets and its ability to quickly converge to a solution, making it a popular choice for initial explorations of energy consumption patterns [10]. Hierarchical clustering, on the other hand, provides a more granular view by constructing a dendrogram that illustrates the relationships among different consumption profiles, thus allowing for a deeper understanding of the data's structure [11]. DBSCAN stands out by effectively identifying clusters of varying densities, which can be particularly beneficial in distinguishing between

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96

normal consumption behaviors and outliers, thereby facilitating targeted interventions [12]. By systematically comparing these three clustering algorithms, this study seeks to determine which method yields the most accurate and actionable insights for optimizing energy consumption in non-residential settings. The findings could contribute to the growing body of literature on energy management and sustainability, highlighting the critical role of advanced data analysis techniques in driving effective energy policies and practices [13], [14].

The novelty of this study lies in its comparative evaluation of clustering performance across two different annual datasets, 2023 and 2024, collected from the same campus environment. Unlike previous research that typically focuses on single-year or static datasets, this approach reveals how the effectiveness of clustering algorithms may vary over time due to evolving consumption behaviors, noise patterns, or operational changes. By highlighting the impact of temporal dynamics on clustering outcomes, this research provides deeper insight into the adaptability of machine learning methods in real-world applications. Ultimately, the insights gained from this comparative analysis empower organizations to make informed decisions that align with their sustainability goals while enhancing overall operational efficiency.

2. METHODOLOGY

The methodology of this study consists of several steps designed to analyze electricity consumption patterns on campus. An overview of the methodology is presented in Figure 1 below.

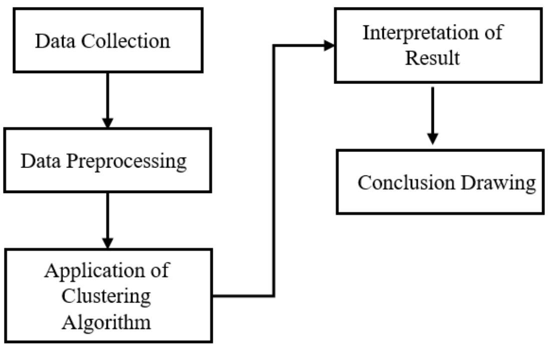

Figure 1. Methodology for this Study

Based on Figure 1, the methodological steps can be described as follows:

2.1. Data Collection

The first step in the methodology involves a thorough data collection process, which is critical for ensuring the accuracy and reliability of the analysis. This process begins with the extraction of information from smart meters installed across various locations on the campus. Smart meters are advanced devices that provide real-time monitoring of electricity usage, allowing for precise tracking of energy consumption at different times of the day and under varying operational conditions.

The timeframe for data collection spans the years 2023 (Dataset 1) and 2024 (Dataset 2), which is strategically chosen to encompass both seasonal variations and potential changes in energy consumption behaviors due to evolving campus activities or policies. By collecting data over this extended period, the study aims to identify trends and patterns that may not be apparent from a shorter timeframe.

To ensure the dataset's integrity, particular attention is given to the accuracy of the data collected. This includes verifying the calibration of smart meters to prevent discrepancies that could arise from faulty devices. Additionally, the data collection process is designed to capture a wide range of variables, such as peak usage times, and overall consumption levels. This comprehensive dataset serves as the foundation for subsequent analyses, enabling researchers to delve into the intricacies of electricity consumption patterns [7], [15].

2.2. Data Preprocessing

The next step is data preprocessing, it is a step in analyzing electricity consumption patterns, ensuring the integrity and reliability of the dataset prior to applying clustering algorithms. This phase begins with data cleaning, where outliers or erroneous entries are removed to enhance the dataset's quality, as these anomalies can skew analysis results [7]. Following this, handling missing values is essential; methods such as interpolation or removal of incomplete entries are employed to maintain robustness [16] [17] [18].

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96

The normalization process further standardizes the dataset, ensuring all features contribute equally to the clustering process. Techniques like Min-Max scaling or Z-score normalization adjust the data to a common scale, which is vital for accurately identifying distinct usage behaviors across various campus facilities [7]. The culmination of these preprocessing steps results in a clean, complete, and normalized dataset that serves as the foundation for subsequent analyses, enhancing the accuracy of the analysis and laying the groundwork for informed decision-making regarding energy management strategies and sustainability initiatives. The formula of the standard scaler is formulated in Formula 1.

$$z = \frac{x - \mu}{\sigma} \qquad (1)$$

In Formula (1), the z score represents the standardized values, x represents the real values, μ represents the meaning of the feature, and σ represents the standard deviation of the dataset. The transformation ensures that the standardized data has a mean of zero and a standard deviation of one. Standardization is crucial in clustering process, as it prevents features with larger numerical ranges from dominating the clustering process.

2.3. Application of Clustering Algorithm

The application of clustering algorithms is a pivotal step in analyzing electricity consumption patterns on campus, as it allows researchers to uncover distinct usage behaviors that can inform energy management strategies. In this study, three prominent clustering techniques: K-means, hierarchical clustering, and DBSCAN are implemented on the cleaned and normalized dataset derived from smart meters and energy monitoring systems. K-means clustering is particularly favored for its efficiency in handling large datasets, quickly converging to a solution that categorizes consumption profiles based on proximity to cluster centroids [10] [19] [20]. This method is adept at identifying general patterns in energy usage, making it a popular choice for initial explorations of consumption data [14]. The advancement of this algorithm can be easily computationally efficient, requiring minimal parameter tuning, is well-suited for large-scale datasets, etc [21].

The formula (2) represents the objective function of the K-Means clustering algorithm, which is commonly used to measure the total intra-cluster distance:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 \qquad (2)$$

J is the total cost or loss function that the algorithm tries to minimize; k is the number of clusters; Ci is the i-th cluster; x represents a data point belonging to cluster Ci; μi is the centroid of cluster Ci; while ‖x−μi‖ represents the Euclidean distance between the data point x and the cluster centroid μi.Hierarchical clustering is one of the unsupervised learning method used to group data based on the similarity constructs a dendrogram that provides a visual representation of the relationships among different consumption profiles, allowing for a more nuanced understanding of the data's structure [6]. The clustering process relies on distance metric such as Euclidean or Manhattan distance [22]. However, this algorithm sensitive to outlier that can reduce the performance of the algorithm [22].This technique is beneficial for identifying subgroups within the dataset that may exhibit similar consumption behaviors, thereby facilitating targeted interventions [22].

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is also an dendogram clustering algorithm that creates a group of data points. It distinguishes itself by effectively identifying clusters of varying densities, which can be particularly advantageous in distinguishing between normal consumption behaviors and outliers [23]. This capability enables stakeholders to focus on specific energy usage patterns that may require immediate attention or intervention [24].
The formula (3) represents the ε-neighborhood of a point x in the context of DBSCAN:

$$N_\varepsilon(x) = \{y \,\varepsilon\, D \; ||x - y|| \;\leq\; \varepsilon\} \qquad (3)$$

N_ε(x) is the ε-neighborhood of point x, i.e., the set of all points y in the dataset D whose distance from x is less than or equal to ε. By applying these three clustering algorithms to the dataset, researchers can conduct a comparative analysis to evaluate the effectiveness of each method in

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96

revealing meaningful patterns in electricity consumption. The insights gained from this analysis not only enhance understanding of energy usage across campus facilities but also support the development of tailored energy management strategies that align with sustainability goals. Ultimately, the application of these advanced clustering techniques serves to empower organizations in making informed decisions that optimize energy consumption and promote responsible usage practices [6].

2.4. Interpretation of Result

After applying K-means, hierarchical clustering, and DBSCAN to the cleaned and normalized dataset, the next step involves a thorough examination of the clusters formed by these algorithms. Each clustering technique reveals distinct electricity usage patterns that can be indicative of various consumption behaviors across different campus facilities. For instance, K-means clustering may highlight general trends in energy usage, such as peak consumption periods during weekdays when classes are in session, whereas hierarchical clustering can provide a more detailed view, illustrating relationships among closely related consumption profiles through a dendrogram [25]. This visual representation aids in identifying subgroups within the dataset that may demonstrate similar consumption behaviors, allowing stakeholders to tailor energy management strategies accordingly. Meanwhile, DBSCAN's ability to discern clusters of varying densities can effectively pinpoint outliers and abnormal consumption events that might warrant immediate attention, such as sudden spikes in energy usage that could signal equipment malfunctions or inefficiencies [26].

This comprehensive analysis provides valuable insights that can inform targeted interventions and energy-saving initiatives, ultimately enhance operational efficiency and align with broader sustainability goals. All the models are evaluated using the Silhouette Score. The Silhouette Score is a metric used to evaluate how well the algorithm clusters the data points. It measures the similarity between the data points in each cluster with the range -1 to 1 [27] [28]. Moreover, the high silhouette score indicates that the data points are similar to their cluster, while the low score indicates a low performance of the clustering algorithm [29].

2.5. Comparative Analysis

In this phase, we conduct a comparative analysis of the clustering results to evaluate the effectiveness of each algorithm in revealing meaningful patterns [30]. The comparison involves assessing each algorithm based on various evaluation metrics such as silhouette score. This metric provides quantitative insights into the quality of clustering results. In addition, the interpretability and practical implications of the resulting clusters are considered to determine which algorithm best uncovers meaningful consumption patterns. Through this comparative process, the strengths and limitations of different algorithms such as K-Means, DBSCAN, and Hierarchical Clustering can be systematically identified. The insights derived are essential in promoting decision-making processes that are informed by empirical data, especially in the development of customized interventions focused on specific user demographics or periods of heightened consumption.

2.6. Conclusion Drawing

After the model development process, the result will be visualized with a visualization diagram. This visualization can be analyzed to indicate the performance algorithm with the value of the Silhouette Score. It is essential to highlight the most effective clustering method for optimizing energy management strategies and informing future sustainability initiatives on campus [20]. This final phase focuses on drawing informed conclusions that summarize the findings and provide strategic recommendations.

3. RESULT AND DISCUSSIONS

3.1. Data Pre-processing

Data pre-processing is a crucial step in ensuring that the dataset is clean, consistent, and suitable for applying algorithms to cluster. In this study, two electricity consumption datasets were used: the 2023 dataset, which consists of 365 rows and 6 columns, and the 2024 dataset, which consists of 366 rows and 6 columns due to the leap year. These datasets contain daily electricity usage values collected from various buildings within the campus environment. Before applying any

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96

machine learning algorithms, especially clustering methods, it is essential to address the presence of missing values, which could otherwise distort analytical outcomes.

Rather than deleting the rows or columns with missing entries, which could result in a loss of significant patterns or introducing bias, this study adopts a more robust approach by inputting missing values using the meaning of each feature [21]. This technique maintains the integrity and completeness of the dataset, ensuring that the patterns and relationships between data points remain intact for accurate clustering.

This approach helps maintain the consistency and completeness of the data, which is crucial for the performance of clustering algorithms. After addressing the missing values, the standard scaler is applied to normalize the dataset. This normalization step is necessary because clustering algorithms, such as K-means and DBSCAN, rely on distance-based calculation and can be significantly affected by features with different scales [23].

Beyond ensuring data consistency, these pre-processing steps has important role in handling the inherent variability and potential noice in electricity consumption data. For instance, fluctuations caused by irregular bulding usage or external environmental condition can create outliers that, if left unstandardized, may dominate cluserting results. By normalizing the data, the clustering process can better reflect the true distribution of consumption accors buildings, rather than beimg skewed by extreme values.

3.2. The Clustering Algorithms Result

In this research, we implement three clustering algorithms, such as K-means, Hierarchical and DBSCAN algorithm for the two datasets. For the K-means algorithm, it is necessary to determine the optimal value of *k*. The value of *k* is identified using elbow method, which results in the same *k* value of 2 for both datasets. This algorithm clusters the data into two groups of low consumption and high consumption. The result is presented in Figure 2 and Figure 3.

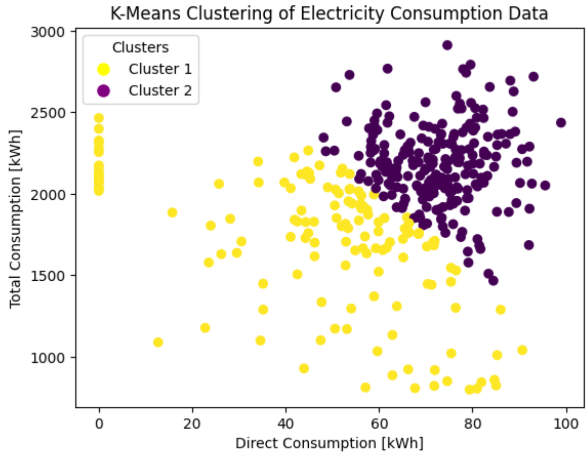

Figure 2. K-means Clustering of Low and High Electricity Consumption in 2023

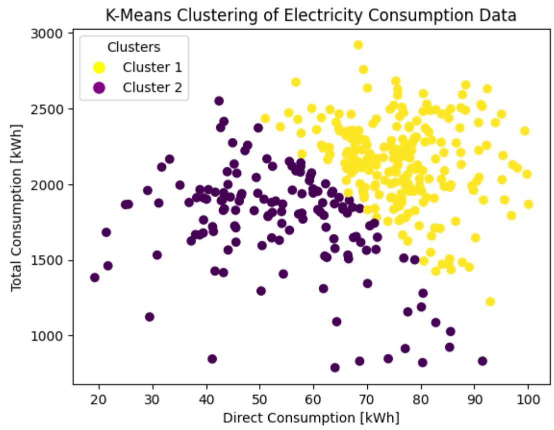

Figure 3. K-means Clustering of Low and High Electricity Consumption in 2024

Figure 2 and Figure 3 present the result of the K-Means clustering applied to the electricity 2023 and 2024 consumption dataset, respectively. The first plot shows the clustering result for 2023 dataset, where the cluster appear less compact and are influenced by a number of outliers with low direct consumption values. In contrast, the second plot illustrates the clustering result for 2024 dataset, which appears more consistent data in the 2024 dataset, which shows more structured and shows better separation between clusters.

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96

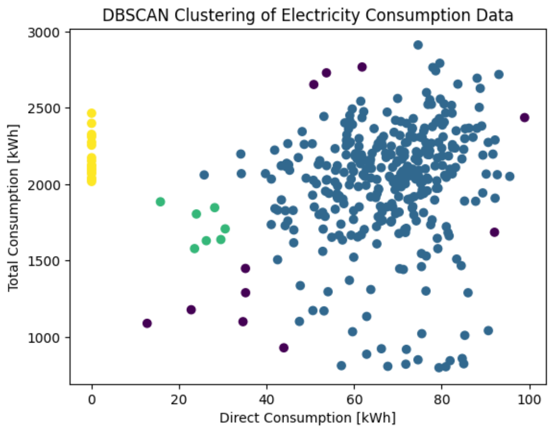

Figure 4. DBSCAN Clustering of Low and High Electricity Consumption in 2023

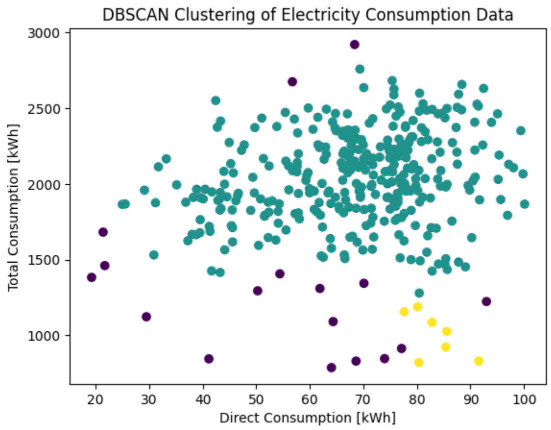

Figure 5. DBSCAN Clustering of Low and High Electricity Consumption in 2024

For the DBSCAN algorithm has different performance for the two datasets. In 2024 electricity dataset, the algorithm can cluster the data into two clusters and 17 noise points presented in Figure 4. This figure illustrates a high direct consumption and relatively low total consumption. The outliers are spread cross lower-density area, suggesting irregular or anomalous consumption patterns. From this result shows the algorithm can captures the structure of the data without assuming predefined cluster shapes, and its ability to detect noise points highlights its strength in identifying a typical behaviour in electricity usage.

For the second dataset, this algorithm can cluster the data into three clusters and 11 noise points illustrated in Figure 5. These algorithm shows a very low direct consumption but high total consumption, possibly due to indirect usage or data anomalies. There is a small cluster that indicate the moderate usage behaviour. This also demonstrated DBSCAN'S advanced in identifying cluster anomalies without assuming any specific cluster shape. This outcome corroborates the findings of the research by Jain et al. [23], who highlighted DBSCAN in identifying both global and local anomalies.

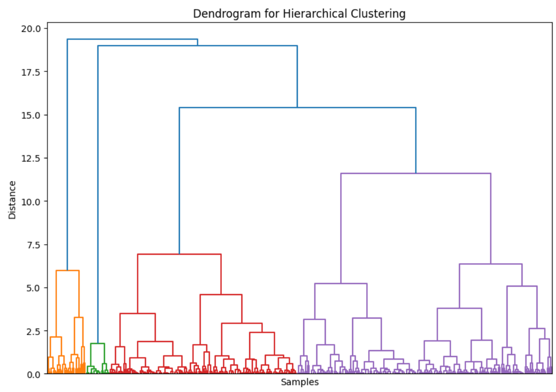

Figure 6. Dendogram for Hierarchical Clustering Electricity Consumption in 2023

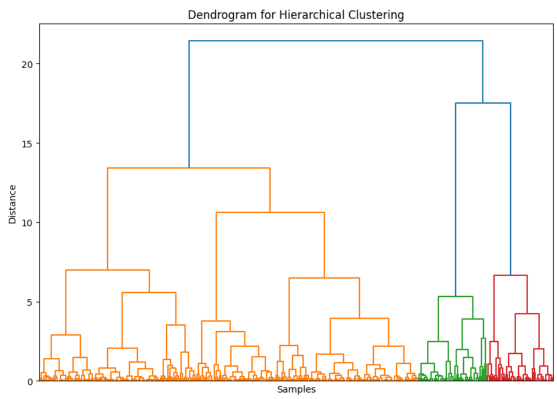

Figure 7. Dendogram for Hierarchical Clustering Electricity Consumption in 2024

The first dendrogram presented in Figure 6 illustrates the outcome of hierarchical clustering utilizing an agglomerative approach, in which data point initially represent individual cluster and is subsequently merged based on a measure similarity. The vertical lines denoted the merging of clusters, and the height at which these merges occur reflects the degree of dissimilarity between the clusters. A substantial increase in linkage distance is observed the range of approximately 15 to 20, indicating a cut this level result in three primary clusters.
In contrast, the 2023 electricity dataset exhibits a more compact hierarchical structure,

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96

characterized by most cluster merges occurring at relatively low dissimilarity values. The branches are more uniformly distributed, and the separations between clusters are less pronounced. A notable division emerges only at the uppermost level of the dendrogram, where two large clusters are joined at a distance slightly above 20, as presented in Figure 7. Applying a cut at this level would result in three clusters. However, the internal composition of these clusters suggests a higher degree of similarity among them when compared to those in the first dendrogram. This indicates that the dataset demonstrates lower inter-cluster variance and reflects a more homogeneous distribution overall. All the results indicate that each dataset has its best-performing algorithm.

As shown in Table 1, the performance of clustering algorithms varies across the two datasets. The highest Silhouette Score is achieved by the Hierarchical Clustering algorithm on Dataset 2. This performance indicates that the algorithm was able to cluster the dataset more effectively. Conversely, the lowest performance is also observed from the same algorithm on Dataset 1, with a Silhouette Score of 0.38. It is also worth noting the fluctuation in the performance of Hierarchical Clustering across datasets. While it yielded the best result in Dataset 2, it performed the worst in Dataset 1. This observation highlights the sensitivity of clustering algorithms to the underlying characteristics of the data. The 2023 dataset may contain more noise or irregularities, making it less suitable for Hierarchical Clustering.

Dataset 2, the best performing algorithm is K-Means, which achieved a silhouette score of 0.44. This indicates that the algorithm was able to generate relatively compact and well separated clusters within the dataset. In contrast, for Dataset 1, the best performing algorithm is DBSCAN, suggesting that is more effective under conditions where outliers are present, as illustrated in Figure 4. This finding aligns with previous research, which has emphasized DBSCAN's strength in detecting anomalies within electricity consumption data [24].

Overall, these findings emphasize that no single clustering algorithm performs best under all conditions. DBSCAN offers greater flexibility and robustness in identifying irregular consumption behavior and outliers, making it suitable for datasets with anomalies[23] . In contrast, K-Means and Hierarchical Clustering tend to perform well when the data exhibits more clearly defined patterns and compact clusters. Therefore, the choice of clustering algorithm should be guided by the specific properties of the dataset, such as distribution, density, and noise level. The table (table number) has a similar evaluation result of the silhouette score, which is 0.53. This result indicates the ability of DBSCAN that does not depend on the cluster center. The performance of all algorithms presented in Table 1.

Table 1. The Clustering Algorithms Performance

Data	Method	SCORE
Dataset 1	K-Means	0.40
	DBSCAN	0.48
	Hierarchical	0.38
Dataset 2	K-Means	0.44
	DBSCAN	0.41
	Hierarchical	0.53

4. CONCLUSION

In this research, we evaluated the effectiveness of three clustering algorithms, i.e., the K-means algorithm, the DBSCAN algorithm, and the Hierarchical algorithm, with electrify usage from Astra Polytechnic in the years 2023 and 2024. The model was evaluated using the Silhouette Score, a widely accepted internal validation metric. The Silhouette Score measures the degree of similarity between an individual data point and its cluster relative to other clusters. The score ranges from -1 to 1, where a higher value indicates that the data points are well clustered, with clear separation between the clusters.

In dataset 1, the best evaluation score was achieved by the DBSCAN algorithm with the value of silhouette score of 0.48. While dataset 2, the best performing algorithm, is achieved by a Hierarchical clustering algorithm with the value of F1-score is 0.53. Therefore, the dataset has its optimal performance clustering algorithm. This suggests that the choice of clustering algorithm should be carefully adapted to the nature of the dataset, particularly

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96

considering factors such as noise, distribution patterns, and the presence of outliers.

REFERENCES

[1] A. Shahcheraghian and A. Ilinca, “Advanced Machine Learning Techniques for Energy Consumption Analysis and Optimization at UBC Campus: Correlations with Meteorological Variables,” *Energies*, vol. 17, no. 18, p. 4714, Sep. 2024, doi: 10.3390/en17184714.

[2] M. Alnahhal, O. Antar, A. Sakhrieh, and M. Al Hazza, “Analyzing Energy Consumption in Universities: A Literature Review,” *Int. J. Energy Econ. Policy*, vol. 14, no. 3, pp. 18–27, May 2024, doi: 10.32479/ijeep.15517.

[3] H. Yesilyurt, Y. Dokuz, and A. S. Dokuz, “Data-driven energy consumption prediction of a university office building using machine learning algorithms,” *Energy*, vol. 310, p. 133242, Nov. 2024, doi: 10.1016/j.energy.2024.133242.

[4] W. Wu, Q. Deng, X. Shan, L. Miao, R. Wang, and Z. Ren, “Short-Term Forecasting of Daily Electricity of Different Campus Building Clusters Based on a Combined Forecasting Model,” *Buildings*, vol. 13, no. 11, p. 2721, Oct. 2023, doi: 10.3390/buildings13112721.

[5] I. B. Sulistiawati, A. Soetedjo, F. Y. Limpraptono, and S. Priyanto, “K-Means Clustering of Electricity Consumption from IoT Data: A Case Study in Electrical Engineering Department Building, ITN Malang,” in *2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Malang, Indonesia: IEEE, Nov. 2023, pp. 434–440. doi: 10.1109/COMNETSAT59769.2023.10420576.

[6] A. M. Alonso, F. J. Nogales, and C. Ruiz, “Hierarchical Clustering for Smart Meter Electricity Loads based on Quantile Autocovariances,” *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4522–4530, Sep. 2020, doi: 10.1109/TSG.2020.2991316.

[7] G. E. Okereke, M. C. Bali, C. N. Okwueze, E. C. Ukekwe, S. C. Echezona, and C. I. Ugwu, “K-means clustering of electricity consumers using time-domain features from smart meter data,” *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, p. 2, Jan. 2023, doi: 10.1186/s43067-023-00068-3.

[8] Assistant Professor, Department of Computer Engineering Institute of Science, Karadeniz Technical University, Turkey, M. Qasimi, A. F. Nasrat, and Master student at Computer science, Gazi Üniversitesi, Turkey., “IoT Sensor Network Electricity Consumption Behaviour Using ClusterAnalysis Algorithm for Network Environment,” *J. Sens. IoT Health Sci.*, vol. 2, no. 3, pp. 46–58, Sep. 2024, doi: 10.69996/jsihs.2024016.

[9] J. Liu, “Cluster Analysis of Electricity Consumption Characteristics and Electricity Consumption Behavior of Typical Industry Users,” in *2023 5th Asia Energy and Electrical Engineering Symposium (AEEES)*, Chengdu, China: IEEE, Mar. 2023, pp. 1389–1395. doi: 10.1109/AEEES56888.2023.10114302.

[10] B. Nepal, M. Yamaha, H. Sahashi, and A. Yokoe, “Analysis of Building Electricity Use Pattern Using K-Means Clustering Algorithm by Determination of Better Initial Centroids and Number of Clusters,” *Energies*, vol. 12, no. 12, p. 2451, Jun. 2019, doi: 10.3390/en12122451.

[11] J. Yang, K. Liu, Z. Deng, M. Zhang, and L. Gong, “Intelligent anomaly detection method for power and electricity data based on cubic exponential smoothing model,” *J. Phys. Conf. Ser.*, vol. 2666, no. 1, p. 012039, Dec. 2023, doi: 10.1088/1742-6596/2666/1/012039.

[12] G. Yao, C. Guo, Q. Ge, and M. Ait-Ahmed, “A practical building energy consumption anomaly detection method based on parameter adaptive setting DBSCAN,” *Cogn. Comput. Syst.*, vol. 3, no. 2, pp. 154–168, Jun. 2021, doi: 10.1049/ccs2.12015.

[13] Y. Cui, Z. Zhu, X. Zhao, and Z. Li, “Energy Schedule Setting Based on Clustering Algorithm and Pattern Recognition for Non-Residential Buildings Electricity Energy Consumption,” *Sustainability*, vol. 15, no. 11, p. 8750, May 2023, doi: 10.3390/su15118750.

[14] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, “K-Means and Alternative Clustering Methods in Modern Power Systems,” *IEEE Access*, vol. 11, pp.

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96

119596–119633, 2023, doi: 10.1109/ACCESS.2023.3327640.
[15] M. Li, Z. Wang, Y. Qu, K. M. Chui, and M. Leung-Shea, “A multi-year campus-level smart meter database,” *Sci. Data*, vol. 11, no. 1, p. 1284, Nov. 2024, doi: 10.1038/s41597-024-04106-1.
[16] J. Chen, J. Yuan, W. Chen, A. Zeb, M. Suzauddola, and Y. A. Nanehkaran, “Research on Interpolation Method for Missing Electricity Consumption Data,” *Comput. Mater. Contin.*, vol. 78, no. 2, pp. 2575–2591, 2024, doi: 10.32604/cmc.2024.048522.
[17] M. Liu *et al.*, “Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques,” *Artif. Intell. Med.*, vol. 142, p. 102587, Aug. 2023, doi: 10.1016/j.artmed.2023.102587.
[18] J. Peng, J. Hahn, and K.-W. Huang, “Handling Missing Values in Information Systems Research: A Review of Methods and Assumptions,” *Inf. Syst. Res.*, vol. 34, no. 1, pp. 5–26, Mar. 2023, doi: 10.1287/isre.2022.1104.
[19] A. A. Abdulnassar and L. R. Nair, “Performance analysis of Kmeans with modified initial centroid selection algorithms and developed Kmeans9+ model,” *Meas. Sens.*, vol. 25, p. 100666, Feb. 2023, doi: 10.1016/j.measen.2023.100666.
[20] S. Mutrofin, T. Wicaksono, and A. Murtadho, “Perbandingan Kinerja Algoritma Kmeans dengan Kmeans Median pada Deteksi Kanker Payudara,” vol. 5, no. 1, 2023.
[21] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *J. Ilm. Inform. Dan Ilmu Komput. JIMA-Ilk.*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.
[22] T. Li, A. Rezaeipanah, and E. M. Tag El Din, “An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3828–3842, Jun. 2022, doi: 10.1016/j.jksuci.2022.04.010.
[23] A. Zahra and N. Asghar, “Clustering Algorithms: An Investigation of K-mean and DBSCAN on Different Datasets,” 2023.
[24] J. Ghanim, M. Issa, and M. Awad, “An Asymmetric Loss with Anomaly Detection LSTM Framework for Power Consumption Prediction,” in *2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON)*, Jun. 2022, pp. 819–824. doi: 10.1109/MELECON53508.2022.9842895.
[25] A. A. Abdullah *et al.*, “Advanced Clustering Techniques for Speech Signal Enhancement: A Review and Metanalysis of Fuzzy C-Means, K-Means, and Kernel Fuzzy C-Means Methods”.
[26] A. K. Abdulsahib, M. A. Balafar, and A. Baradarani, “DGBPSO-DBSCAN: An Optimized Clustering Technique Based on Supervised/Unsupervised Text Representation,” *IEEE Access*, vol. 12, pp. 110798–110812, 2024, doi: 10.1109/ACCESS.2024.3440518.
[27] G. Vardakas, I. Papakostas, and A. Likas, “Deep Clustering Using the Soft Silhouette Score: Towards Compact and Well-Separated Clusters,” Feb. 01, 2024, *arXiv*: arXiv:2402.00608. doi: 10.48550/arXiv.2402.00608.
[28] V. Michalakopoulos, E. Sarmas, I. Papias, P. Skaloumpakas, V. Marinakis, and H. Doukas, “A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs,” *Appl. Energy*, vol. 361, p. 122943, May 2024, doi: 10.1016/j.apenergy.2024.122943.
[29] A. Hilmani, Y. Sabri, A. Maizate, S. Aouad, and M. Koundi, “Integration of K-Means and Silhouette score for energy efficiency of wireless sensor networks,” *Int. J. Reconfigurable Embed. Syst. IJRES*, vol. 14, no. 1, p. 26, Mar. 2025, doi: 10.11591/ijres.v14.i1.pp26-34.
[30] S. K. A. Ashgar, L. Gungat, A. M. Sarman, and N. Bolong, “An Evaluation of Sustainable Campus Guidelines in Energy Management Context”.

DOI : https://doi.org/10.21063/jtif.2025.V13.2.87-96