JOSAR, Vol.10 No.2 September 2025
ISSN: 2502-8521 (Print) / 2503-1155 (Online)
DOI: https://doi.org/10.35457/josar.v10i2.5232

Comparative Study Of Monthly Electricity Consumption Clusterization Using K-Means and DBSCAN

Putri Rochfiani[1]
Mukh Taofik Chulkamdi[2]
Udkhiati Mawaddah[3]

[1,2,3]Informatics Engineering, Engineering and Informatics, Balitar Islamic University, Blitar, STUDENT

Abstract

Analyzing electricity consumption patterns is essential for improving the efficiency of energy distribution and identifying anomalies such as unusual spikes or possible electricity theft. This research presents a comparative analysis of two clustering algorithms, K-Means and DBSCAN in classifying monthly electricity usage of customers from PT PLN (Persero) Rayon Ngunut, covering the districts of Rejotangan, Ngunut, Kalidawir, and Pucanglaban. The dataset includes consumption records from November and December 2024. The K-Means algorithm, which uses a centroid-based clustering approach, performs effectively on uniform datasets, while DBSCAN, a density-based method, is more capable of recognizing outliers and non-spherical cluster formations. The performance of both algorithms was evaluated using Accuracy, Mean Squared Error (MSE), Precision, Recall, and F1-Score. Experimental outcomes revealed that K-Means achieved an accuracy of 96%, MSE of 0.0400, precision of 0.71, recall of 1.00, and an F1-score of 0.83. In contrast, DBSCAN reached an accuracy of 76%, MSE of 0.2400, precision of 0.29, recall of 1.00, and an F1-score of 0.45. These results demonstrate that K-Means produces more compact and consistent clusters, while DBSCAN is superior in identifying anomalies, detecting 17 outliers in total. Consequently, K-Means is considered more suitable for stable consumption grouping, whereas DBSCAN is recommended for anomaly detection purposes. The findings are expected to assist PT PLN (Persero) in developing data-driven and adaptive strategies for more efficient energy management.

Keywords: Clustering, DBSCAN, Electricity consumption, K-Means, PLN

1. INTRODUCTION

Electricity has become an essential aspect of modern human life, supporting industrial, economic, and domestic activities Akbar, (2023) . The increasing dependence on electrical energy requires efficient management to ensure its stability and sustainability. One of the challenges faced by electricity providers such as PT PLN

[1]Corresponding author, email: rochfianisunarto@gmail.com

Citation in APA style: Rochfiani, P, Chulkamdi, M, Mawaddah, U. (2025). Comparative Study Of Monthly Electricity Consumption Clusterization Using K-Means and DBSCAN. *JOSAR,* Vol.10. No.(2):221-229.

Received:	**Revised:**	**Published:**
October, 06st 2025	October, 13st 2025	October, 14st 2025

ISSN: 2502-8521 (Print) / 2503-1155 (Online)
DOI: https://doi.org/10.35457/josar.v10i2.5232

(Persero) is understanding customer consumption behavior to support effective energy distribution and identify potential anomalies such as abnormal usage or electricity theft, based on customer electricity consumption data obtained from PT PLN Rayon Ngunut (2024). Clustering techniques play a crucial role in classifying electricity usage patterns by identifying similarities among data points. Through the implementation of clustering algorithms, customers with comparable consumption behaviors can be organized into particular groups that support effective monitoring, forecasting, and managerial decision-making. Among the numerous clustering approaches Gustriansyah, Suhandi, & Antony, (2019) .

The K-Means algorithm operates based on centroids, dividing data into several clusters according to the minimum distance between data points and their nearest cluster center Indriani, Irawan, & Bahtiar, (2024) . This method performs efficiently on uniform datasets and generates distinct cluster boundaries. Conversely, DBSCAN applies a density-based principle, where data points that share close proximity within dense regions are grouped together, while those located in sparse regions are classified as noise or outliers. This characteristic enables DBSCAN to adapt effectively when dealing with datasets that exhibit irregular distributions Mutiah et al., (2024).

Previous studies have compared these algorithms in various domains. Research by Adha, Nurhaliza, Sholeha, & Mustakim, (2021) demonstrated that K-Means outperformed DBSCAN in terms of cluster validity with a Silhouette Index of 0.6902 compared to 0.3624. Similarly, Akbar (2023) found that K-Means produced more consistent clusters with a Silhouette Score of 0.845, while Mutiah, Hasnataeni, Fitrianto, & Jumansyah, (2024) concluded that DBSCAN was more effective in detecting outliers. These findings suggest that algorithm performance varies depending on data characteristics and research objectives.

The purpose of this research is to conduct a comparative analysis between the K-Means and DBSCAN algorithms in clustering monthly electricity consumption data obtained from PT PLN (Persero) Rayon Ngunut, which includes four districts Rejotangan, Ngunut, Kalidawir, and Pucanglaban. The dataset utilized in this study contains electricity usage records for the months of November and December 2024. To assess and compare algorithm performance, several evaluation indicators were employed, namely The performance evaluation employed several metrics, including Mean Squared Error (MSE), Accuracy, Precision, Recall, and the F1-Score, to measure the effectiveness of each clustering algorithm. The findings of this analysis are anticipated to offer a deeper understanding of the advantages and limitations of each algorithm, as well as to support PT PLN (Persero) in enhancing data-driven decision-making processes aimed at achieving more efficient and adaptive electricity distribution management.

2. LITERATURE REVIEW

2.1. K-Means Algorithm

K-Means is a clustering technique designed to group datasets into several clusters based on the similarity among data points. The main objective of this algorithm is to minimize the overall squared distance between each observation and the centroid of its corresponding cluster. The procedure starts by defining the number of clusters (k) and initializing the centroids at random positions. Subsequently, each data point is allocated to the nearest centroid according to the Euclidean distance measure. Once all data points have been assigned, the centroid locations are recalculated as the mean of the points within each cluster. This iterative process continues until the centroids no longer change significantly or the specified maximum number of iterations has been completed (Kuo et al., 2005).According to Priati & Fauzi, (2017), this iterative process ensures that the algorithm produces optimal groupings. The K-Means method is efficient for large datasets and can clearly separate data into specific groups. Nur Afidah, (2023) applied K-Means to analyze population migration data in Rembang Regency, dividing the data into three clusters high, medium, and low migration levels showing that K-Means effectively identifies population movement patterns.

2.2 DBSCAN Algorithm

DBSCAN is a density-oriented clustering technique that identifies clusters by locating regions with a high concentration of data points and distinguishing them from areas of lower density. Points that are located far from dense regions are treated as outliers. This algorithm relies on two essential parameters: Epsilon (ε), which specifies the greatest allowable distance between two samples for them to be regarded as neighbors, and MinPts, which defines the minimum number of nearby points needed to establish a dense region. According to these criteria, each data instance is classified as a core, border, or noise point, depending on its proximity and relationship to surrounding data within the spatial structure Mutiah et al., (2024).

2.3 Previous Studies

Several previous studies have compared the performance of K-Means and DBSCAN algorithms in various application domains. **Akbar (2023)** conducted a study on clustering electricity consumption patterns in Madrasah Aliyah schools in East Java and reported that the K-Means algorithm achieved a higher Silhouette Score (0.845) compared to DBSCAN (0.825), indicating better structure and compactness. Similarly, **Adha et al. (2021)** analyzed global COVID-19 data and found that K-Means produced a higher Silhouette Index (0.6902) than DBSCAN (0.3624), demonstrating that K-Means performs better on structured datasets. In contrast, **Qadrini (2020)** applied both algorithms to laboratory performance data at ITS and found that DBSCAN outperformed K-Means with a Silhouette Coefficient of 0.72 versus 0.40, suggesting DBSCAN's superiority for datasets with varying density and noise. Meanwhile, **Andriyani and Puspitarani (2022)** reported that DBSCAN achieved slightly higher accuracy (99.80%) than K-Means (99.50%) in clustering product review data, although DBSCAN required a longer processing time.

3. METHODS

This study applied a quantitative descriptive approach to compare the performance of the K-Means and DBSCAN algorithms in clustering monthly electricity consumption data Sulianta & Widyatama, (2024). The purpose of this stage was to analyze the effectiveness of both algorithms in forming groups that represent electricity consumption behavior among customers.

The dataset was obtained from PT PLN (Persero) Rayon Ngunut, which includes four districts: Rejotangan, Ngunut, Kalidawir, and Pucanglaban. The data consist of monthly electricity usage for November and December 2024. Each record contains information such as customer identification, district, and total electricity consumption in kilowatt-hours (kWh).

ISSN: 2502-8521 (Print) / 2503-1155 (Online)
DOI: https://doi.org/10.35457/josar.v10i2.5232

Before the clustering process, several preprocessing steps were carried out to ensure the quality and consistency of the data. The steps included removing duplicate and incomplete records and normalizing the data so that each attribute had the same contribution in the analysis. The DBSCAN algorithm required an additional process to determine the optimal Epsilon (ε) and MinPts values based on the characteristics of the dataset through a k-distance approach.

Both algorithms were then applied to the processed dataset. K-Means divided the data into three clusters representing low, medium, and high electricity consumption levels, while DBSCAN grouped data points based on density and identified outliers that did not belong to any cluster. The results from both methods were evaluated using accuracy, mean squared error (MSE), precision, recall, and F1-score to determine which algorithm provided better clustering performance.

4. RESULTS

4.1 K-Means Clustering Results

The K-Means algorithm successfully grouped electricity consumption data into three clusters representing low, medium, and high usage levels. The clustering process converged after several iterations and produced compact and well-separated clusters. Most data points were distributed within the medium cluster, indicating that the majority of customers have stable electricity usage.

The evaluation results of K-Means performance are shown in Table 1. The algorithm achieved an accuracy of 96%, The results showed that the algorithm achieved, MSE value of 0.0400, with precision, recall, and F1-score recorded at 0.71, 1.00, and 0.83, respectively. These results indicate that the K-Means algorithm provides consistent and accurate cluster formation with minimal error.

Table 1. K-Means Clustering Performance Results

Evaluation Metric	Value
Accuracy	96 %
Mean Squared Error (MSE)	0.0400
Precision	0.71
Recall	1.00
F1-Score	0.83

Table 1 shows that K-Means produces high clustering accuracy with a low MSE value, indicating stable and consistent performance. The algorithm is highly reliable for identifying electricity consumption patterns with minimal classification error.

The percentage of each cluster formed by K-Means is illustrated in Table 2. It shows that the medium consumption group dominates the dataset, followed by low and high consumption groups.

Table 2. K-Means Cluster Distribution

Cluster	Category	Percentage
Cluster 0	Low Consumption	24.0 %
Cluster 1	Medium Consumption	56.0 %
Cluster 2	High Consumption	20.0 %

Table 2 shows that most customers fall into the medium consumption category (56%), followed by low (24%) and high (20%) categories. This indicates that electricity usage in the four districts is relatively moderate and stable.

ISSN: 2502-8521 (Print) / 2503-1155 (Online)
DOI: https://doi.org/10.35457/josar.v10i2.5232

4.2 DBSCAN Clustering Results

The DBSCAN algorithm grouped data points based on their density and identified points in sparse regions as noise or outliers. The clustering process resulted in fewer clusters than K-Means and detected 17 outliers, representing abnormal electricity usage.

As shown in Table 3, DBSCAN achieved an accuracy of 76%, a Mean Squared Error (MSE) of 0.2400, a precision of 0.29, a recall of 1.00, and an F1-score of 0.45. Compared to K-Means, the overall clustering performance of DBSCAN was lower, but it provided better sensitivity in identifying anomalous data.

Table 3. DBSCAN Clustering Performance Results

Evaluation Metric	**Value**
Accuracy	76 %
Mean Squared Error	0.2400
Precision	0.29
Recall	1.00
F1-Score	0.45

Table 3 shows that DBSCAN achieved lower accuracy compared to K-Means but demonstrated higher sensitivity in detecting noise and anomalies. The high recall value indicates that the algorithm effectively recognized most relevant data points despite its lower precision.

The results indicate that DBSCAN tends to form irregular cluster shapes due to variations in data density. Table 4 shows the percentage distribution of the formed clusters.

Table 4. DBSCAN Cluster Distribution

Cluster	**Category**	**Percentage**
Cluster 0	Normal Consumption	68.0 %
Cluster 1	High Consumption	32.0 %
Outliers	-	17 data points

Table 4 shows that normal electricity consumption dominates the dataset (68%), while high consumption accounts for 32%. Additionally, 17 data points were detected as outliers, representing abnormal usage patterns. This confirms that DBSCAN is effective in identifying irregularities in electricity consumption data.

5. DISCUSSION

The comparison between the application of K-Means and DBSCAN techniques for clustering analysis of monthly electricity consumption data shows significant differences in performance, data grouping characteristics, and anomaly detection. The results reveal that the K-Means algorithm produces more structured and stable clusters, while the DBSCAN algorithm demonstrates higher sensitivity to data density variations and performs well in identifying abnormal electricity usage behavior. This finding aligns with the theory proposed by Farid, (2024), which states that K-Means performs effectively on datasets with uniform distribution, whereas DBSCAN is more suitable for datasets with arbitrary shapes and varying densities.

Based on the evaluation metrics, K-Means achieved an accuracy of 96%, a Mean Squared Error (MSE) of 0.0400, a precision of 0.71, a recall of 1.00, and an F1-score of 0.83. These results indicate that K-Means has strong clustering consistency and can accurately separate customer groups according to their electricity consumption levels.

ISSN: 2502-8521 (Print) / 2503-1155 (Online)
DOI: https://doi.org/10.35457/josar.v10i2.5232

The low MSE value also reflects that the distance between data points and their respective cluster centroids is minimal, signifying that the formed clusters are compact and homogeneous. This result supports the findings of Akbar, (2023) and Adha et al., (2021), who both reported that K-Means outperformed DBSCAN in forming compact and well-defined clusters in structured datasets such as electricity usage and COVID-19 data, respectively.

In contrast, the DBSCAN algorithm obtained an accuracy of 76%, an MSE of 0.2400, a precision of 0.29, a recall of 1.00, and an F1-score of 0.45. Although its overall accuracy is lower than that of K-Means, DBSCAN successfully detected 17 outliers, which represent irregular or inconsistent electricity consumption data. This result is consistent with Qadrini, (2020)**,** who found that DBSCAN performs better than K-Means when dealing with datasets that contain noise or variable densities. The ability of DBSCAN to detect anomalies also supports the original concept by Sufairoh et al., (2023) who introduced DBSCAN as a density-based clustering algorithm capable of identifying outliers effectively without the need to predefine the number of clusters.

From the clustering results, K-Means formed three clusters consisting of low consumption (24%), medium consumption (56%), and high consumption (20%). The dominance of the medium consumption category indicates that most customers under PT PLN Rayon Ngunut tend to have stable electricity usage. The low consumption cluster generally consists of customers with minimal activity or efficient energy use, while the high consumption cluster includes customers with greater energy demand, such as households with more appliances or small business operations. This interpretation is in line with Andriyani & Puspitarani, (2022), who also found that K-Means is effective in grouping consumer data with regular and consistent patterns.

Meanwhile, DBSCAN formed two major clusters, namely normal consumption (68%) and high consumption (32%), and identified 17 outliers. These outliers represent data that deviate significantly from the average usage pattern. The existence of these anomalies may indicate possible electricity theft, technical recording errors, or sudden increases in electricity load. This finding aligns with the study by Qadrini (2020), which highlighted DBSCAN's strength in identifying irregularities and noise within real-world datasets.

Overall, the analysis results confirm that the K-Means algorithm is superior in producing well-defined and consistent clusters when the data distribution is regular, while DBSCAN excels in detecting anomalies and irregular consumption patterns in datasets with varying densities. Both algorithms have their respective strengths and can complement each other when applied together. The combination of K-Means and DBSCAN can therefore be a more comprehensive approach for PLN to understand customer consumption behavior K-Means for identifying general usage patterns and DBSCAN for detecting unusual or abnormal data.

In practical terms, these findings can assist PT PLN (Persero) Rayon Ngunut in monitoring electricity consumption patterns across different regions, improving energy distribution planning, and identifying customers who display irregular consumption trends. The comparative analysis also provides a reference for future studies on clustering-based anomaly detection in electricity consumption, particularly in optimizing customer segmentation and improving the accuracy of predictive energy management systems.

ISSN: 2502-8521 (Print) / 2503-1155 (Online)
DOI: https://doi.org/10.35457/josar.v10i2.5232

6. CONCLUSION

This study compared the performance of the K-Means and DBSCAN algorithms in clustering monthly electricity consumption data from PT PLN (Persero) Rayon Ngunut, which includes the districts of Rejotangan, Ngunut, Kalidawir, and Pucanglaban. The research aimed to determine the most effective algorithm for grouping electricity consumption behavior and identifying anomalies or irregular usage patterns.

The results show that the K-Means algorithm produced better performance in terms of clustering accuracy and compactness, with an accuracy of 96%, MSE of 0.0400, precision of 0.71, recall of 1.00, and F1-score of 0.83. K-Means was able to form three well-separated clusters representing low, medium, and high consumption categories. The majority of customers were classified in the medium consumption group, indicating that electricity usage in the four districts tends to be stable and moderate.

In contrast, the DBSCAN algorithm achieved lower clustering accuracy with an accuracy of 76%, MSE of 0.2400, precision of 0.29, recall of 1.00, and F1-score of 0.45, but it successfully detected 17 outliers representing abnormal electricity usage patterns. This finding shows that DBSCAN is more capable of identifying anomalies and irregular consumption behavior, which is essential for detecting potential electricity losses or data inconsistencies.

Overall, the study concludes that K-Means performs better for datasets with homogeneous characteristics and stable patterns, while DBSCAN is more suitable for detecting anomalies in data with variable density. The combination of both algorithms can provide a comprehensive understanding of electricity consumption behavior — where K-Means identifies general consumption groups, and DBSCAN detects unusual usage patterns.

The findings of this research are expected to support PT PLN (Persero) in optimizing electricity distribution strategies, improving monitoring systems, and developing data-driven decision-making models. Future studies can further enhance this analysis by including additional parameters such as customer type, seasonal consumption variations, or geographic factors to obtain more detailed insights into electricity usage behavior.

REFERENCES

Adha, R., Nurhaliza, N., Sholeha, U., & Mustakim, M. (2021). Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, *18*(2), 206–211. https://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/12469

Akbar, E. (2023). Perbandingan Algoritma Dbscan-K Means dan K Means untuk Pengelompokan Madrasah Aliyah Provinsi Jawa Timur. *Repository.Uinjkt.Ac.Id*, 32–51. https://repository.uinjkt.ac.id/dspace/handle/123456789/68812%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/68812/1/EMIR AKBAR-FST.pdf

ISSN: 2502-8521 (Print) / 2503-1155 (Online)
DOI: https://doi.org/10.35457/josar.v10i2.5232

Andriyani, F., & Puspitarani, Y. (2022). Performance Comparison of K-Means and DBScan Algorithms for Text Clustering Product Reviews. *SinkrOn*, *7*(3), 944–949. https://doi.org/10.33395/sinkron.v7i3.11569

Farid, M. (2024). *Menggunakan Algoritma Density Based Spatial Clustering of Application With Noise (Dbscan) Tugas Akhir*.

Gustriansyah, R., Suhandi, N., & Antony, F. (2019). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science*, *18*(1), 470–477. https://doi.org/10.11591/ijeecs.v18.i1.pp470-477

Indriani, D., Irawan, B., & Bahtiar, A. (2024). Penerapan Algoritma K-Means Clustering Untuk Menentukan Persediaan Stok Barang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 182–187. https://doi.org/10.36040/jati.v8i1.8322

Kuo, R. J., Wang, H. S., Hu, T. L., & Chou, S. H. (2005). Application of ant K-means on clustering analysis. *Computers and Mathematics with Applications*, *50*(10–12), 1709–1724. https://doi.org/10.1016/j.camwa.2005.05.009

Mutiah, S., Hasnataeni, Y., Fitrianto, A., & Jumansyah, L. M. R. D. (2024). *Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat Dalam era digital saat ini , jumlah data yang tersedia dari berbagai bidang , termasuk sosial dan ekonomi , terus*. *09*(September), 247–260.

Nur Afidah, N. (2023). Penerapan Metode Clustering dengan Algoritma K-means untuk Pengelompokkan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang. *PRISMA, Prosiding Seminar Nasional Matematika*, *6*, 729–738. https://journal.unnes.ac.id/sju/index.php/prisma/

Priati, & Fauzi, A. (2017). Data Mining dengan Teknik Clustering Menggunakan Algoritma K-Means pada Data Transaksi Superstore. *Seminar Nasional Informatika Dan Aplikasinya (SNIA)*, *September*, hal. 15-19. http://community.tableau.com.

Qadrini, L. (2020). Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, *13*(2), 5–11. https://doi.org/10.36456/jstat.vol13.no2.a2886

Sufairoh, I., Rani, A. C., Amalia, K., & Rolliawati, D. (2023). Perbandingan Hasil Analisis Clustering Metode K-Means, DBSCAN Dan Hierarchical Pada Data Marketplace Electronic Phone. *JOINS (Journal of Information System)*, *8*(1), 97–105. https://doi.org/10.33633/joins.v8i1.8016

Sulianta, F., & Widyatama, U. (2024). *Metode K-Means Clustering dalam Pengelompokan Penjualan Produk Indofood Metode K-Means Cluster ing dalam Pengelompokan Penjualan Produk Indofood Fakultas Teknik , Universitas Widyatama Bandung , Indonesia maupun goreng , susu , bumbu tambahan untuk memasa*. *March*.