Pandemic Impact on Residential Electricity Customer Load Profile Clustering

Maricar M. Navarro
Industrial Engineering Department
Technological Institute of the Philippines
Cubao, Quezon City
mnavarro.ie@tip.edu.ph

Bryan B. Navarro
Electrical Engineering Department
Technological Institute of the Philippines
Cubao, Quezon City
bryan.navarro@tip.edu.ph

Abstract

This study focuses on analyzing customer load profiles and their impact on the electric distribution grid in the context of the COVID-19 pandemic in the Philippines. With the increasing number of Filipinos working from home, energy consumption has surged, making it crucial for electric distribution companies to classify customer profiles based on energy usage. In this paper, we collected residential data from a Philippine electric distribution company during the COVID-19 pandemic. The objective is to employ clustering methods, particularly the K-means algorithm, to group customer profiles and determine the optimal number of clusters using the multi-criteria model and elbow method. All simulation were performed using MATLAB platform. This study provides valuable insights for electric distribution companies in managing customer load profiles and optimizing the distribution grid amidst the pandemic.

Keywords
Clustering, Customer Load Profile, Elbow Method, K-means, Pandemic

1. Introduction
Electric distribution companies in the Philippines play a critical role in providing reliable and efficient electricity services to consumers. To optimize operations and ensure grid stability, these companies are keenly interested in analyzing customer load profiles. Understanding the energy consumption patterns of customers and classifying them into distinct categories allows for better demand forecasting, resource allocation, and grid management strategies. However, the emergence of the COVID-19 pandemic has introduced new challenges and uncertainties, significantly impacting the energy consumption landscape and requiring a reevaluation of customer load profiles.

A customer load profile refers to a comprehensive representation of the energy consumption patterns exhibited by individual customers over a specific period. It provides valuable insights into the variations in energy usage throughout the day, week, month, or year. Load profiles are essential tools in energy management, demand response programs, load forecasting, and overall grid optimization. By analyzing customer load profiles, utilities, grid operators, and energy service providers can gain a deeper understanding of consumption patterns, identify energy-saving opportunities, and make informed decisions about energy generation, distribution, and pricing strategies, (Yang et al. 2018)

Customer load profiles are typically constructed using interval metering data, which records the energy consumption at regular intervals, such as every 15 minutes or every hour. These data points capture the dynamic nature of energy usage, allowing for a granular analysis of consumption patterns. Load profiles can reveal valuable information about

peak demand periods, seasonal variations, and specific usage patterns associated with different customer segments, (Patel et al. 2019).

The analysis of customer load profiles plays a vital role in energy planning and grid management. Utilities can utilize load profiling to determine the capacity requirements for distribution networks, optimize load balancing, and identify potential areas of congestion or voltage issues. Load profiles also assist in identifying energy efficiency opportunities and demand response strategies. By understanding how customers use energy and when peak demand occurs, utilities can design targeted programs to incentivize load shifting or energy conservation during peak periods, ultimately reducing strain on the grid and lowering costs for both customers and utilities.

Load profile clustering is a popular technique used to group customers with similar energy usage patterns. Clustering algorithms, such as the K-means algorithm, help identify distinct clusters or segments of customers based on their load profiles. This enables utilities to tailor their offerings, such as personalized energy efficiency recommendations or demand response programs, to specific customer groups. Load profile clustering also facilitates anomaly detection, allowing utilities to identify abnormal energy consumption patterns that may indicate equipment malfunctions, metering errors, or even unauthorized energy usage (Sharma and Chilamkurti 2019).

The COVID-19 pandemic has had far-reaching effects on the daily lives of people around the world, including the Philippines. As the government implemented various measures to control the spread of the virus, such as lockdowns and social distancing guidelines, people's routines and habits changed drastically. Work-from-home arrangements became prevalent, and businesses faced disruptions, leading to a significant shift in energy consumption patterns.

In the context of the electric distribution sector, this shift in energy consumption has been particularly notable. With more people staying at home, the demand for residential electricity has surged. Traditional patterns of energy consumption, where peak demand was primarily driven by commercial and industrial sectors during working hours, have been upended. The increased reliance on home offices, online learning, and indoor activities has resulted in higher energy consumption during daytime hours and altered load profiles.

Against this backdrop, electric distribution companies in the Philippines face the challenge of adapting to and managing these changes effectively. It becomes crucial to reevaluate customer load profiles in light of the pandemic's impact to develop strategies that optimize grid operations, ensure reliable service delivery, and promote sustainability.

1.1 Objectives
This study focuses on examining the impact of the COVID-19 pandemic on customer load profile clustering in a Philippine electric distribution company. A comprehensive dataset comprising 1,153,433customer profiles from the year 2022 is analyzed to understand the evolving energy consumption patterns during this exceptional period. By utilizing clustering methods, particularly the K-means algorithm, the study aims to categorize customer profiles based on their energy usage characteristics and identify distinct clusters representing specific consumption patterns. In addition to clustering, this paper aims to determine the optimal number of clusters. The multi-criteria model and elbow method are employed as validation techniques to ensure the robustness of the clustering results. The multi-criteria model considers various factors such as intra-cluster homogeneity and inter-cluster separation, providing a comprehensive evaluation of the clustering quality. The elbow method, on the other hand, examines the within-cluster sum of squares to identify the point where adding more clusters does not significantly improve the clustering performance.

2. Related Literature
Clustering is a data mining approach used to divide data points into groups or clusters based on their similarity (Kumar & Goyal, 2013). It is commonly applied in various fields such as image processing, pattern recognition, market research, and data analysis. One of the popular clustering techniques is the K-means algorithm. It is an unsupervised learning technique used when the data is unlabeled, meaning there are no predefined categories or groups. The algorithm works iteratively to assign data points to one of the K clusters based on their features or characteristics. The similarity of data points is measured using a distance metric, often the Euclidean distance. The algorithm aims to minimize the within-cluster sum of squares, which represents the squared distance between each data point and the centroid of its assigned cluster.

The K-means algorithm is widely used in various industries, especially when dealing with large amounts of unlabeled data. It can help discover hidden patterns or groups within the data, which can be valuable for businesses. By clustering customers based on their purchasing behavior, for example, companies can identify new customer segments or target specific groups with tailored marketing strategies. In the context of load profile clustering, the K-means algorithm has been utilized in research studies. Load profile clustering involves grouping energy consumption profiles based on similarities in usage patterns. This can be beneficial for energy management, demand response, or anomaly detection. By applying the K-means algorithm to load profiles, different clusters of energy usage patterns can be identified, enabling better understanding and analysis of energy consumption data.

It is worth noting that the K-means algorithm has certain limitations. It requires the number of clusters (K) to be specified in advance, which can be a challenge when the optimal number of clusters is unknown. Additionally, the algorithm is sensitive to the initial placement of cluster centroids, which may result in different solutions. Various extensions and variations of the K-means algorithm have been proposed to address these limitations and improve clustering performance.

Several studies have examined the impact of the COVID-19 pandemic on energy consumption patterns and the application of clustering techniques in customer load profile analysis. Silva et al. (2021) and Tijerina et al. (2020) have highlighted the significant increase in residential electricity consumption due to work-from-home arrangements and changes in peak demand periods. Guo et al. (2020) and Liu et al. (2019) have utilized the K-means algorithm to classify customer load profiles into distinct patterns representing different consumer behaviors.

 They demonstrated the effectiveness of clustering techniques in understanding energy consumption patterns and optimizing grid operations. Li et al. (2020) employed clustering algorithms to analyze customer load profiles based on different temporal resolutions, while Kabir et al. (2020) and Zhang et al. (2021) explored the application of hierarchical clustering and fuzzy C-means in uncovering hidden patterns in energy consumption data. Kumar and Kamal (2020) proposed the use of the elbow method to determine the optimal number of clusters, while Ibrahim et al. (2019) and Chen et al. (2020) utilized multi-criteria models for validating clustering results. Liu et al. (2021) and Ochoa et al. (2018) focused on the electric distribution industry and demonstrated the value of customer load profile clustering in optimizing grid management and developing targeted strategies for demand-side management. However, a research gap exists in the combined analysis of customer load profile clustering and the impact of the COVID-19 pandemic in the context of the Philippine electric distribution industry. This study aims to address this gap by conducting a case study in a Philippine electric distribution company, analyzing customer load profiles amidst the pandemic, and employing the K-means algorithm for clustering, while also validating the number of clusters using the multi-criteria model and elbow method.

3. Methodology
The methodology employed in this study, focuses on analyzing the impact of the pandemic on customer load profile clustering using the K-Means algorithm. The methodology consists of the following key processes:

1. **Data Preparation***:*
    Data Collection: historical customer load profile data from the Philippine Electric Distribution Company, including attributes such as customer ID, and energy consumption
    Data Cleansing: Remove duplicates, handle missing values, and address any data inconsistencies to ensure data quality.
    Data Normalization: Scale the load profile data to a common range to ensure fair comparisons during the clustering process.

2. K-Means Clustering and Evaluation:
    Initialization: Randomly initialize the cluster centroids, starting with a minimum of two clusters.
    Iterative Clustering: Apply the K-Means algorithm iteratively, gradually increasing the number of clusters until the maximum specified number (20 in this study) is reached.
    Evaluation: Utilize multi-criteria evaluation method and elbow method, to determine the optimal number of clusters. Simulate K-Means clustering again using the chosen number of clusters to obtain cluster centroids and labels for each data point.

3. Cluster Profile Analysis:

Cluster Characteristics: Analyze and examine the characteristics of each cluster, considering load profiles and behaviors during different phases of the pandemic.

Notable Changes: Identify any significant changes or trends within clusters, such as load reductions, shifts in peak demand, or altered load profiles.

4. Optimal Cluster Determination:

Determine the Optimal Cluster: Use the results of multi-criteria evaluation methods (e.g., elbow method) to select the optimal number of clusters that best represent the data's underlying structure and capture the impact of the pandemic on customer load profiles.

5. Documentation and Analysis:

Analysis and Insights: Interpret the findings and insights derived from the cluster analysis, highlighting the impact of the pandemic on customer load profiles. Provide recommendations or suggestions based on the results obtained.

K-Means

1. *Input:* Number of clusters (K)

2. *Initialization*: Generate K initial cluster centroids randomly or using more advanced techniques like k-means++.

3. *Assignment:* For each data point, calculate the distance (e.g., Euclidean, Manhattan) to each centroid.

Assign the data point to the nearest centroid, forming K clusters.

4. *Re-computation:* Update the centroid of each cluster by calculating the mean or median of the data points assigned to that cluster.

5. *Convergence Check:* Calculate the differences between the old and new centroids of each cluster. If the maximum difference across all clusters is lower than a pre-defined tolerance limit, then stop. Otherwise, proceed to the next iteration.

6. *Iterative Refinement*: Return to the Assignment Step and reassign each data point to the nearest centroid based on the updated centroids. Repeat the Re-computation and Convergence Check steps until convergence is achieved.

7. *Finalization:* The algorithm terminates when convergence is reached or a maximum number of iterations is reached.

Validity Index Assessment

1. Multi-criteria Method

Different clustering strategies applied to the same dataset can produce varying results due to the different approaches employed. To assess the reliability of the resulting clustering solution, this paper proposes a multi-criteria model that incorporates widely recognized validity indices, including the Silhouette Score (SI), Calinski-Harabasz Score (CH), and Davies-Bouldin Score (DB).

$$MC = w_1 S_{norm} + w_2 C_{norm} + w_3 (1 - D_{norm})$$
(1)

$$w_1 + w_2 + w_3 = 100\%$$
(2)

Where;

MC = multi-criteria value

Snorm, Cnorm, and Dnorm = normalized score index

w1, w2, w3 = weights

The multi-criteria model for cluster determination involves the following steps:

1.1 Silhouette Score (S))

The Silhouette Score (SI Score) is an evaluation indicator that measures how similar a data point is within a cluster (cohesion) to other clusters (separation). As shown in (4), the ultimate average value obtained s(i) overall points in a dataset with a specified cluster centers *k*, (Tambunan et al, 2020)

$$S_{SCORE} = \max_{k} s(k)$$
(3)

In (4), the silhouette value of a single data point i specifies how strongly the data sets in a certain cluster are grouped.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4}$$

Where;

    a= average distance between i and all the other data points in the cluster to which i belongs
    b= average distance between i and all clusters to which i does not belong

The silhouette value s(i) can range from -1 to 1, as shown in (6), but when s(i) is close to 0, a(i) and b(i) have nearly equal value.

$$1 \leq s(i) \leq 1 \tag{5}$$

A high value implies the object matches well with its cluster and not very well with the surrounding clusters; a low value indicates that the object does not match the adjacent clusters poorly. Clustering settings are helpful when the majority of objects are of high value. If the value of many points is low or negative, there may be too many or too few clusters in the clustering settings.

1.2 Calinski-Harabasz Score (C)

Calinski and Harabasz proposed the Calinski-Harabasz Score (C Score), which may be used to evaluate the model when ground truth labels are unknown and validate how successfully the clustering has been done using dataset quantities and features (Chou et al, 2004)

$$C_{SCORE} = \frac{\sum_{k=1}^{K} n_k \times \|C_k - C\|^2}{\sum_{k=1}^{K} \sum_{i=1}^{n_k} \|X_{ik} - C_k\|^2} \times \frac{N - K}{K - 1} \tag{6}$$

Where;
n_k = The number of points of the kth cluster
CK = the centroid of the kth cluster
c = global centroid
N = overall data points

Although there is no "good" cut-off number for the C Score, a higher value indicates that the clusters are dense and well divided. We must choose the solution that produces a peak or at least a sharp elbow on the line plot of C scores.

1. Calculate the Silhouette Score and Calinski-Harabasz Score for each clustering solution. The clustering solution with the highest Silhouette Score and Calinski-Harabasz Score is considered the best in terms of compactness and separation, indicating better clustering quality.
2. Select the clustering solution that yields the highest values for the Silhouette Score and Calinski-Harabasz Score as the optimal number of clusters.

1.3 Davies-Bouldin Score (D)
The Davies-Bouldin Score (D Score) is then simply the average of all clusters' maximum values, as determined by the equation in (8), (Davies and Bouldin, 1979). *(D)* assesses each cluster's quality by comparing it to other clusters and then offers the average case for all clusters. It is abbreviated as DB, and these constraints require that the Index be symmetric and non-negative. This is calculated as a function of volatility within the cluster, and the ratio of separation between clusters, so smaller values indicate better clustering.

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \tag{7}$$

$$D_{SCORE} = \frac{1}{K} \sum_{i=1}^{K} D_i \qquad (8)$$

Where;
Rij = metric to determine how effective a clustering technique is
Mij = separation between i-th and j-th clusters (ideally large)
Si = the cluster spread of cluster i (Ideally it should be as low as possible)

Both the data and the algorithm determine this. Di selects the worst-case scenario equal to Rij for the most similar cluster to cluster i. This formulation could take many different forms, such as using the cluster similarity average, a weighted average, etc.

1. Compute the Davies-Bouldin Score for each clustering solution. The clustering solution with the lowest Davies-Bouldin Score indicates better cluster separation and cohesion, reflecting higher quality clustering.
2. Choose the clustering solution that results in the lowest Davies-Bouldin Score as the optimal number of clusters. By considering the Silhouette Score, Calinski-Harabasz Score, and Davies-Bouldin Score, the multi-criteria model provides a comprehensive evaluation of the clustering solutions. The model selects the clustering solution with the highest Silhouette and Calinski-Harabasz Scores, indicating better compactness and separation, while also favoring the solution with the lowest Davies-Bouldin Score, indicating improved cluster cohesion and separation

2. Elbow Method
The Elbow Method serves as the second set of cluster determination to evaluate the ideal number of clusters (k) for a given assignment. It offers a valuable graphical tool to assess the optimal value of *k (*Tan et al, 2018). By applying the Elbow Method, the optimal number of clusters is determined by identifying the elbow point on the SSE curve. This method provides a visual means to understand the impact of increasing the number of clusters on the SSE and aids in finding the most appropriate value of *k* for the given dataset and clustering task.

4. Results and Discussion
The effectiveness of the proposed method is demonstrated using real residential data collected from an electric distribution company in the Philippines during the COVID-19 pandemic. The dataset comprises 1,153,433 customer profiles from the year 2022. During the data pre-processing and cleaning stage, 643,468 customers were excluded due to insufficient data and missing variables, resulting in a total of 509,964customers. All simulations were conducted in the MATLAB programming environment. Table 1 presents the summary scores of the three indices.

4.1 Numerical Results

Table 1. Index Score

k cluster	Silhouette Score (Maximum Score)	Calinski-Harabasz Score (Maximum Score)	Davies-Bouldin Score (Minimum Score)
2	1.00	0.00	0.13
3	0.52	0.00	0.33
4	0.46	0.07	0.23
5	0.21	0.15	0.23
6	0.22	0.22	0.04
7	0.17	0.31	0.04
8	0.11	0.42	0.00
9	0.11	0.45	0.34
10	0.05	0.52	0.49
11	0.06	0.56	0.48
12	0.06	0.65	0.53
13	0.06	0.72	0.45
14	0.04	0.78	0.36
15	0.03	0.81	0.58
16	0.04	0.87	0.63
17	0.02	0.90	0.76
18	0.01	0.93	0.84
19	0.01	0.98	1.00
20	0.00	1.00	0.98

The individual criteria assessment indicates that the suggested number of clusters using the Silhouette (S) Score, Calinski-Harabasz (C) Score, and Davies-Bouldin (D) Score was 2, 20, and 8 clusters, respectively .
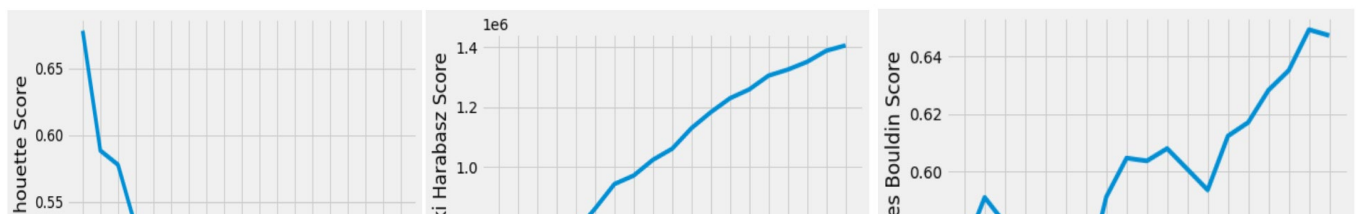

Figure 1. Individual Score

4.1 Optimal Cluster Identification
Optimal cluster identification refers to the process of determining the most appropriate or ideal number of clusters for a given dataset or clustering problem. It involves selecting the number of clusters that best captures the inherent patterns, structures, or groups within the data, leading to meaningful and interpretable results.

4.1.1 Multi-criteria Method
For the multi-criteria method, this paper proposes a weighted approach to maximize the cluster evaluation. The weighted multi-criteria model, outlined below, aims to optimize the clustering solution:

Table 2. Multi-criteria method

Case	w1 (S)	w2 (C)	w3 (D)	No. of Clusters
1	100%	0%	0%	2
2	0%	100%	0%	20
3	0%	0%	100%	8
4	33%	33%	33%	2
5	30%	30%	40%	2
6	30%	40%	30%	2
7	40%	30%	30%	2
8	10%	0	90%	8
9	20%	0	80%	2
10	30%	0	70%	2
11	40%	0	60%	2
12	50%	0	50%	2
13	60%	0	40%	2
14	70%	0	30%	2
15	80%	0	20%	2
16	90%	0	10%	2
17	0	10%	90%	8
18	0	20%	80%	8
19	0	30%	70%	8
20	0	40%	60%	8
21	0	50%	50%	8
22	0	60%	40%	14
23	0	70%	30%	14
24	0	80%	20%	20
25	0	90%	10%	20
26	10%	90%	0	20
27	20%	80%	0	20
28	30%	70%	0	20
29	40%	60%	0	20
30	50%	50%	0	20
31	60%	40%	0	2
32	70%	30%	0	2
33	80%	20%	0	2
34	90%	10%	0	2

Table 3. Multi-criteria method

Cluster Results	Case Count	%
2	17	50%
20	8	24%
8	7	21%
14	2	6%

Based on the data provided in Table 2, the multi-criteria method identified the best number of clusters as 2 clusters, which accounted for 50% of the total number of cases. The summary result for the 34 cases is presented in Table 2, while the detailed results for each case are in Table 3

4.1.2 Elbow Method
This paper employs the "kneel locator" function in the elbow method to precisely determine the number of clusters. By analyzing the resulting plot, the elbow point is identified at k = 7, which is considered an optimal choice for this dataset.

4.1.3 Cluster Profile
Once the optimal number of clusters was determined, the K-means algorithm was simulated again to calculate the cluster centroid and assign cluster labels to each data point. Figure 2. showcase the cluster load profiles, visually representing the energy consumption patterns within each cluster.

In order to effectively illustrate the time series electricity profile, this paper employed a conversion technique from actual consumption usage, measured in kilowatts (kW), to per unit value. This conversion was necessary because the actual values exhibited significant variations, making it challenging to visualize and compare them directly. The per

unit value was obtained by calculating the ratio of the actual consumption value to the average value. This normalization approach helped to standardize the data and bring the values within a comparable range, enabling a more meaningful analysis of the time series electricity profiles. By converting the actual values to per unit values, the paper ensured that the differences between data points were more easily discernible and allowed for better visualization and interpretation of the electricity consumption patterns.
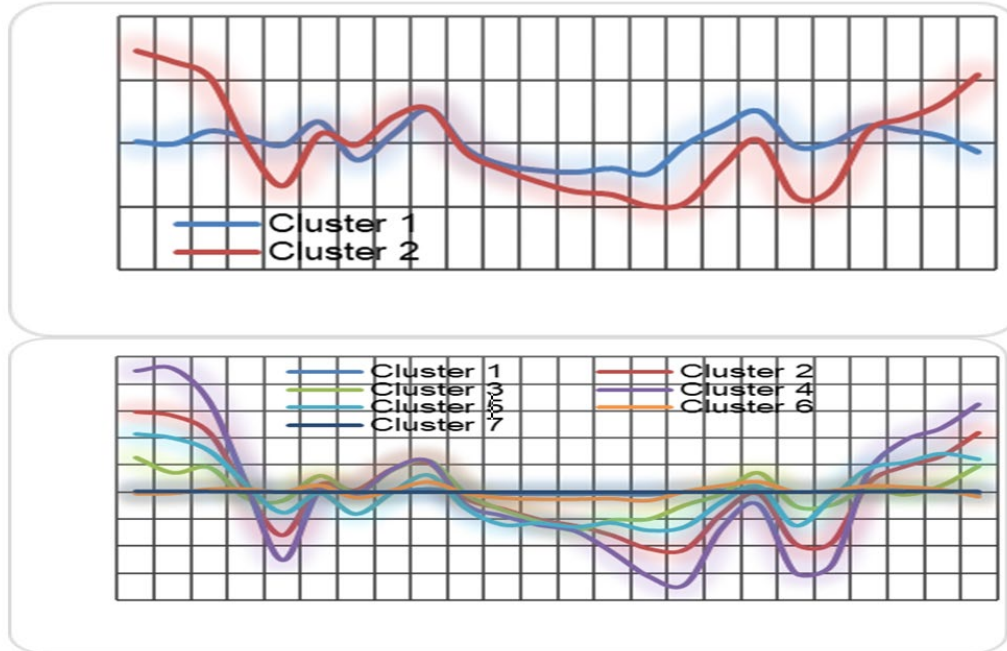

Figure 2. Residential Customer Load Profiles in Multi Criteria (2 clusters) and Elbow Method (7 clusters)

Table 4. Cluster Statistics (Multi-Criteria Method)

Cluster #	Customer Count	%
1	422,311	83%
2	87,654	17%

Table 5. Cluster Statistics (Elbow Method)

Cluster #	Customer Count	%
1	188,197	37%
2	37,263	7%
3	85,720	17%
4	13,911	3%
5	101	0%
6	182,872	36%
7	1,902	0%

From Fig. 2, it can be observed that clusters 1-5 represent customers with similar consumption patterns characterized by higher activity primarily during nighttime hours (8pm - 1am), while exhibiting lower consumption levels during the afternoon (12nn-4pm) and early evening (6pm-7pm) periods. On the other hand, clusters 6-7 exhibit nearly flat consumption patterns, indicating consistent and relatively stable levels of energy usage throughout the day.

5. Conclusion

This paper focuses on the clustering of customers' electricity load profiles using the K-means algorithm. A dataset of approximately 1 million customers was collected from a distribution company in the Philippines. The study utilized the Multi-Criteria Model and Elbow Method to determine the optimal number of clusters, resulting in 2 and 7 clusters, respectively.

The Multi-Criteria Model revealed two distinct peaks in the load profiles: Cluster 1 exhibited higher activity in the morning (8am) and afternoon (5pm), while Cluster 2 showed higher activity in the morning (8am) and evening (12mn). On the other hand, the Elbow Method identified 7 clusters, with Clusters 1-5 demonstrating similar consumption patterns with higher activity at nighttime (8pm-1am) and lower consumption during the afternoon (12nn-4pm and 6pm-7pm). Clusters 6-7 exhibited relatively flat consumption patterns.

The COVID-19 pandemic has had a significant impact on customer load profiles. The analysis of load profiles during the pandemic period revealed shifts in consumption patterns and changes in overall energy demand. These changes can be attributed to various factors such as remote work arrangements, lifestyle adjustments, and shifts in commercial activities. In addition, the pandemic highlighted the importance of adapting load profile clustering techniques to capture and analyze the evolving consumption patterns during exceptional circumstances. Electric distribution companies should consider incorporating temporal and contextual factors into their clustering models to accurately reflect the changes in customer behavior during such events. Moreover, The study underscores the importance of building resilience and flexibility into the electric grid infrastructure. Understanding the changes in load profiles can help companies better anticipate and manage shifts in energy demand. This can inform decisions related to capacity planning, network upgrades, and resource allocation to ensure reliable and efficient energy supply during challenging times. Finally, the pandemic serves as a learning opportunity for electric distribution companies to enhance their preparedness for future crises or disruptive events. By leveraging advanced clustering algorithms, real-time data analytics, and scenario modeling, companies can develop strategies to mitigate the impact of similar situations in the future, ensuring the resilience and reliability of the energy system. These findings provide valuable insights for electric distribution companies, enabling them to establish rules and strategies related to demand response, network planning, and load forecasting.

For future research, it is recommended to explore the combination of the K-means algorithm with Dynamic Time Warping (DTW) for analyzing customer load profiles in time series data. Additionally, incorporating meta-heuristic optimization techniques such as Genetic Algorithm (GA) could enhance the K-means algorithm's ability to determine the optimal number of clusters.

References

Chen, Y., Qian, K., & Yang, F. Multi-objective clustering validation for customer load profile analysis in smart grid." *Energies*, vol. 13, no. 3, pp. 579 , 2020

Chou, C.-H.; Su, M.-C.; Lai, E. A new cluster validity measure and its application to image compression, *Pattern Analysis and Applications* , vol. 7 , pp. 205–220, 2004

Davies, D. L.; Bouldin, D. W. A Cluster Separation Measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (*TPAMI*) vol 2, pp. 224-227, 1979,

Guo, S., Zhang, T., & Zeng, X. Research on customer load clustering based on K-means algorithm. *Energy Procedia,* vol 173, pp. 26-31, 2020

Ibrahim, R., Imran, S., Salim, N., Yahaya, N., & Sulaiman, M. Evaluation and comparison of clustering validation indices *2019 6th International Conference on Electrical and Electronics Engineering (ICEEE)* (pp. 10-15). IEEE., 2019

Kabir, M. I., Shen, Y., & Bokhari, A. A fuzzy C-means based clustering approach for residential load profiling*. In 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm),* pp. 1-6. 2020

Kumar, S., & Kamal, T. S. Determining the optimal number of clusters in K-means clustering using elbow method. *International Journal of Computer Science and Mobile Computing*, vol 9, no. 6, pp. 106-112, 2020

Kumar, M. and Goyal, A. Clustering: A Data Mining Approach for Intrusion Detection. *International Journal of Computer Applications*, vol. 67, no. 16, 2013

Li, H., Zhang, L., Wu, J., & Wu, Z. Multi-time scale customer load profile clustering analysis using K-means algorithm. *Energies* ,vol 13,no.19 ,pp 5201, 2020

Liu, Y., Yao, Z., Xu, D., Wei, Q., & Zhu, Y. Residential electricity load profile clustering based on K-means algorithm. *International Journal of Electrical Power & Energy Systems*, vol 134, 2021

Liu, Y., Zhang, W., & Huang, Y. Clustering analysis of customer load profile based on K-means algorithm. *2019 IEEE Innovative Smart Grid Technologies—Asia (ISGT Asia)* pp. 720-724, 2019

Ochoa, L. F., Ponce, P., Nieva, J., & Mancarella, P. Clustering-based consumer segmentation for demand-side management in distribution networks. *Applied Energy*, vol. 228, pp. 1035-1047, 2018

Patel, A., Kundu, P., Mishra, S., & Roy, M. Load Profiling and Its Applications in Energy Planning and Grid Management: A Review. *International Journal of Electrical Power & Energy Systems*, vol. 114, 2019

Sharma, S., & Chilamkurti, S. Load Profile Clustering for Improved Energy Management: A Review. *Energies*, Vol. 12, No. 19, 2019

Silva, V., Mello, B., Gomes, G., Barreto, R., & Castro, C. Analysis of the impact of COVID-19 on the residential electricity consumption in Brazil. *Energy Reports*, vol 7, pp. 482-489, 2021

Tan, P., Steinbach, M., Karpatne, A., & Kumar, V. Introduction to Data Mining (2nd Edition) (2nd ed.). Pearson. 2018

Tambunan, H. B.; Barus, D. H.; Hartono, J.; Alam, A. S.; Nugraha, D. A.; Usman, H. H. H. Electrical Peak Load Clustering Analysis Using K-Means Algorithm and Silhouette Coefficient, *2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*, pp. 23-24, 2020

Tijerina, L. O., Batista, M. A. V., & Riquelme, F. P. (2020). COVID-19 impact on electricity load forecasting.. *Electric Power Systems Research*, vol. 189, 2020

Yang, H., Dong, J., Wen, Z., He, Z., Gan, D., Cai, M., & Wang, H. Analysis of Customer Load Profiles for Residential Electricity Demand Forecasting: A Review. *Energies*, vol. 11, no. 7, 2018

Zhang, H., Zhang, Y., Huang, B., Chen, D., & Fang, X. Short-term residential load forecasting using hierarchical clustering and long short-term memory networks. *Applied Energy,*vol. 290, 2021

Biography:

Maricar M. Navarro is a Professional Industrial Engineer (PIE) awarded by the Philippine Institute of Industrial Engineers (PIIE) and an ASEAN Engineer (AE) awarded by the ASEAN Federation of Engineering Organizations. She is an Assistant Professor IV in the Department of Industrial Engineering and a Professor of the Graduate School Program at the Technological Institute of the Philippines-Quezon City. Engr. Navarro has done research projects that deal with the optimization of production, warehouse operations, and service operations. Her research interests include manufacturing, simulation, optimization, facility layout, and design, She is an active member and Professional Industrial Engineer of the Philippine Institute of Industrial Engineers (PIIE) organization in the Philippines.

Bryan B. Navarro is an Assistant Professor in the Department of Electrical Engineering at the Technological Institute of the Philippines and is currently a Staff Engineer at Manila Electric Company (MERALCO). He earned B.S. in Electrical Engineering from the Technological Institute of the Philippines, Quezon City, Master of Science in Electrical Engineering major in Power Systems from the University of the Philippines, Diliman, Quezon City. He has published journal and conference papers. He is an active member of the Institute of Integrated Electrical Engineers of the Philippines (IIEE).