engineering proceedings

Proceeding Paper

Prosumer Clustering for Optimized Control and Peer-to-Peer Energy Trading in Solar-PV and Electric Vehicle Integrated Community Microgrids: A Comparative Analysis of K-Means and Spectral Methods †

Mukovhe Ratshitanga [1,2,*], Komla Agbenyo Folly [1] and David Oyedokun [1]

[1] Advanced Electric Power and Energy Conversion (AEPEC) Research Center, Department of Electrical Engineering, University of Cape Town, Rondebosch 7700, South Africa; komla.folly@uct.ac.za (K.A.F.); david.oyedokun@uct.ac.za (D.O.)
[2] Department of Electrical, Electronic and Computer Engineering, Cape Peninsula University of Technology, Bellville 7535, South Africa
* Correspondence: rtsmuk001@myuct.ac.za or ratshitangam@cput.ac.za
† Presented at the 34th Southern African Universities Power Engineering Conference (SAUPEC 2026), Durban, South Africa, 30 June–1 July 2026.

Abstract

This study presents a comprehensive clustering analysis of residential prosumer profiles for optimizing control and peer-to-peer (P2P) energy trading in community renewable energy systems (CRES). Using data from 25 prosumer households equipped with rooftop solar photovoltaic (PV) systems and electric vehicle (EV) charging capabilities, this study implements and compares k-means and spectral clustering algorithms to identify optimal segmentation strategies for prosumer energy management. K-means clustering identifies seven practical prosumer categories with a silhouette coefficient of 0.17, while spectral clustering achieves superior mathematical separation with a silhouette coefficient of 0.275 in ten clusters, though producing six singleton outliers. The k-means solution demonstrates three primary prosumer categories: net producers, net consumers, and balanced profiles. Cluster size variation requires adaptive optimization, while singleton outliers need custom strategies. EV ownership impact consumption, so future proliferation demands dynamic clustering, and these findings will guide metaheuristic algorithms for energy trading and pricing.

Keywords: community renewable energy; data analysis; k-means clustering; microgrids; P2P energy trading; prosumer profile; spectral clustering

Check for updates

Academic Editor: Akshay Kumar Saha

Published: 13 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The increasing availability of high-resolution data and analytical methods has made the investigation of residential electricity consumption dynamics at finely resolved timescales progressively more feasible. Distributed generation and load profile clustering constitutes a prevalent methodological approach to exploring production and consumption dynamics. Notwithstanding the abundance of available algorithmic techniques, clustering load profiles poses challenges, as clustering methods do not invariably capture the temporal aspects of electricity production or consumption, and clusters are difficult to explicate without supplementary descriptive household data. These challenges circumscribe the utility of cluster analysis in elucidating behavioral and other drivers of electricity usage

patterns [1]. The proliferation of smart meters, decentralized energy resources, and sensors has driven the importance of data analysis, leading to increased complexity in processing and leveraging data [2]. Clustering techniques, such as k-means and spectral clustering, can extract underlying patterns from energy consumption data, but determining the most appropriate clusters of consumers with similar behaviours remains a challenge [3,4].

This paper presents an unsupervised machine learning framework for optimal load and prosumer generation profiling, utilizing real consumption data from 25 households. Various clustering algorithms are reviewed and compared, with empirical analysis and evaluation metrics assessing their performance.

2. Literature Review

The integration of solar generation into power systems, particularly at low-voltage and medium-voltage levels, poses significant operational challenges for distribution systems. Centralized voltage control schemes have been proposed to mitigate voltage violations, but their efficacy relies on accurate forecasted data and a reliable communication infrastructure [5]. Research on clustering algorithms has focused on load consumption, with limited attention paid to embedded generation production or surplus energy sharing. Studies have applied clustering approaches to group households by energy consumption and production patterns, enhancing efficiency through dimensionality reduction and feature extraction [6]. However, these approaches often neglect the dynamic nature of surplus energy in communities with rooftop systems. Existing research has also explored clustering algorithms for electricity customer segmentation, aiming to identify groups with similar consumption patterns [7,8]. Techniques such as k-means, fuzzy k-means, and hybrid approaches have been applied, but researchers often overlook fairness and equity issues in community surplus energy sharing. Comparative analyses of clustering algorithms, including spectral, genetic, and adaptive algorithms, have been conducted using real datasets [9]. However, these studies often prioritize market efficiency over community resiliency, misaligning with distributed energy sharing objectives.

Other approaches, such as Gaussian Mixture Models (GMMs) and Self-Organizing Maps (SOMs), have been applied to analyze prosumer and consumer patterns [10,11]. However, these methods can create ambiguity in P2P trading and may not adapt to varying patterns of variability in electrical demand profiles. Table 1 shows the selection of research completed on several load consumption and surplus generation clustering algorithms.

Table 1. Models for clustering residential loads and embedded generation.

Algorithm	Load Clustering	Generation Clustering
K-means	[7–9,12–17]	[10]
Hierarchical clustering	[10,11]	[10]
DBSCAN	[9]	[9]
Genetic algorithms	[9]	[9]
Spectral clustering	[9]	[9]
Gaussian Mixture Model	[10,18]	[10]
Fuzzy K-means	[8]	Lack of clustering in distributed home energy production
Fuzzy C-means	[11,15]	
Self-Organizing Maps	[11,13]	
Artificial Neural Networks	[12]	
Probabilistic Neural Networks	[8]	

Static clustering methods in microgrid energy management have limitations, as they cannot adapt to dynamic changes in load profiles and generation patterns, leading to inefficiencies. Renewable energy variability makes real-time forecasting and demand response optimization complex. Key gaps include uncertainty around ideal cluster size, limited

understanding of effective configurations, and a lack of comprehensive clustering studies. No established methods exist for the simultaneous allocation of distributed generators, EV charging stations, and protection equipment, which are crucial for smart grid reliability. Addressing these gaps can enhance microgrid efficiency, reliability, and adaptability, supporting renewable energy integration and smart grid development.

3. Methodology

This study tackles the challenge of characterizing prosumer energy profiles in community renewable energy management systems, focusing on peer-to-peer energy trading frameworks. A comprehensive clustering analysis was conducted on 25 residential prosumer profiles, incorporating solar photovoltaic generation, household consumption patterns, and electric vehicle charging behaviors. K-means and spectral clustering algorithms were employed to identify optimal prosumer segmentation strategies, using the elbow method and silhouette coefficient as validation metrics. The clustering framework aims to establish homogeneous prosumer groups that facilitate efficient energy trading mechanisms while maintaining computational tractability for real-time optimization algorithms.

The data used for this research, presented in Figure 1, comprise measurements from 25 houses with varying characteristics, including solar rooftop PV embedded generation with storage, electric vehicles, and general household loads [19]. The figure provides insight into similarity analysis. Notably, a total of six days exhibit surplus generation below approximately 3.11 kW. The performances of k-means and spectral clustering algorithms are compared, with descriptions of each method provided in Section 4. This study's findings will enable efficient energy trading mechanisms and inform real-time optimization algorithms.

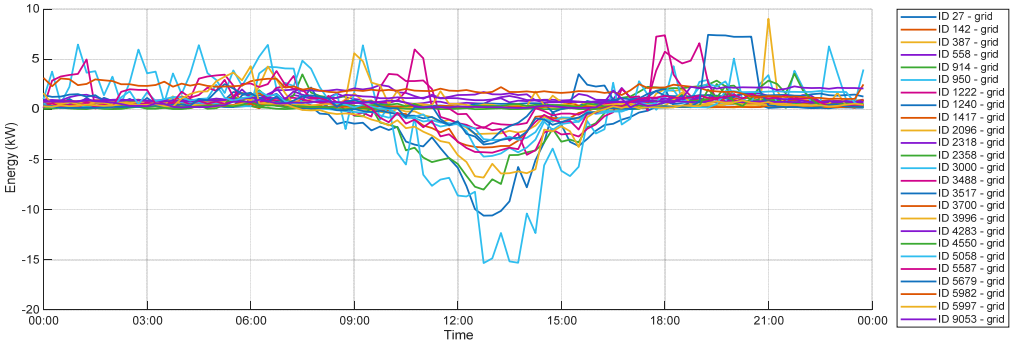


Figure 1. Random day 15-min resolution energy data—25 households.

4. Clustering Algorithms

Clustering algorithms partition data into meaningful groups based on similarity measures, with applications spanning machine learning, data mining, and pattern recognition [20]. Each algorithm addresses specific challenges, from handling different data types, discovering arbitrary cluster shapes and scaling, to massive datasets or adapting to evolving data streams.

4.1. Partitioning Methods: Optimizing Cluster Assignments Through Iterative Refinement—K-Means Clustering

K-means clustering aims to partition n data points into k clusters by minimizing the within-cluster sum of squares distances [21,22]. The algorithm rests on the principle that each data point should belong to the cluster whose centroid (mean) is nearest to it. The theoretical foundation involves minimizing an objective function representing total within-cluster variance, making it suitable for discovering compact, spherical clusters in numerical data.

Objective Function:
The k-means algorithm seeks to minimize:

$$Z(C_1, C_2 \ldots, C_k) = \sum_{\ell=1}^{k} \sum_{i \in C_\ell} \|x_i - \mu_\ell\|_2^2 \quad (1)$$

$$\mu_\ell = \frac{1}{|C_\ell|} \sum_{i \in C_\ell} x_i \quad (2)$$

where C_1 *to* C_k is the set of points in a cluster ℓ, μ_ℓ is the centroid of points in the cluster ℓ, x_i is the i^{th} object of the dataset, and $\|x_i - \mu_\ell\|_2$ denotes the Euclidean norm or distance between the vectors x_i and μ_ℓ. k is the cluster number and C_ℓ is the ℓ^{th} cluster.

4.2. Spectral Clustering: Graph Partitioning Through Eigenvalue Decomposition

Spectral clustering approaches data clustering from a graph partitioning perspective, representing data points as vertices in a weighted similarity graph. This method partitions the graph to group similar points together, leveraging the spectral properties of graph Laplacian matrices to identify arbitrarily shaped clusters. The graph Laplacian encodes connectivity structure, with eigenvectors revealing underlying cluster structure. Spectral clustering relates to random walks on graphs, where good clusters correspond to regions with extended walk durations. The normalized cut (Ncut) criterion balances between-cluster similarity minimization and within-cluster similarity maximization, finding partitions with rare transitions between clusters. While computing exact minimum Ncut is NP-hard, spectral methods provide efficient relaxations via eigenvector solutions [23,24].

4.2.1. Similarity Graph Construction

Given data points $x_1, \ldots, x_n$, construct weighted adjacency matrix W where:

$$w_{ij} = s(x_i, x_j) = e^{\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)} \quad (3)$$

Degree matrix D is diagonal with:

$$d_i = \sum_j w_{ij} \quad (4)$$

where w_{ij} is the similarity weight between datapoints i and j, and $s(x_i, x_j)$ is the similarity function between points x_i and x_j. Datapoints i and j in the original d-dimensional feature space are denoted by x_i and x_j, $\|x_i - x_j\|^2$ is the squared Euclidean norm or distance between x_i and x_j, and the bandwidth parameter of the Gaussian kernel controlling the decay rate of similarities with distance is σ^2. The exponential function creates the Gaussian kernel.

4.2.2. Graph Laplacian Matrices

- Unnormalized Laplacian: $L = D - W$.
- Normalized Random Walk Laplacian: $L_{rw} = D^{-1}L = I - D^{-1}W$.
- Symmetric Normalized Laplacian:

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \quad (5)$$

where I is the identity matrix of an appropriate dimension.

https://doi.org/10.3390/engproc2026140009

4.2.3. Normalized Cut Objective

$$Ncut(A_1, \ldots, A_k) = \frac{1}{2}\sum_j \left[\frac{W(A_i, \overline{A_i})}{vol(A_i)} \right] \quad (6)$$

where:

$$cut(A, B) = \sum_{i \in A, j \in B} W(i, j) \quad (7)$$

and,

$$vol(A) = \sum_{i \in A} d_i \quad (8)$$

where $A_1, \ldots, A_k$ are the k clusters, $W(A_i, \overline{A_i})$ is the total weight of edges between cluster A_i and its complement, and $vol(A_i)$ is the volume of the cluster A_i.

5. Simulation Results

The comparative analysis of clustering methodologies reveals distinct segmentation characteristics that significantly impact the design of community energy management strategies. K-means clustering identified an optimal configuration of seven clusters (silhouette coefficient = 0.17), demonstrating a 47% reduction in within-cluster sum of squares from k = 1 to k = 7, with cluster sizes ranging from 1 to 7 profiles. Figures 2 and 3 show the graphical representation of the solutions.

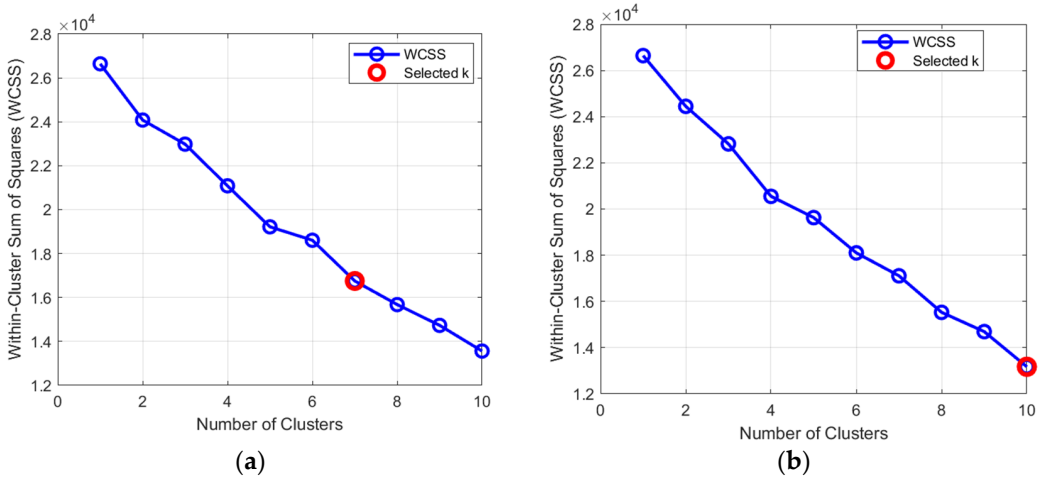


Figure 2. Clustering elbow method comparison for: (**a**) k-means; (**b**) spectral.

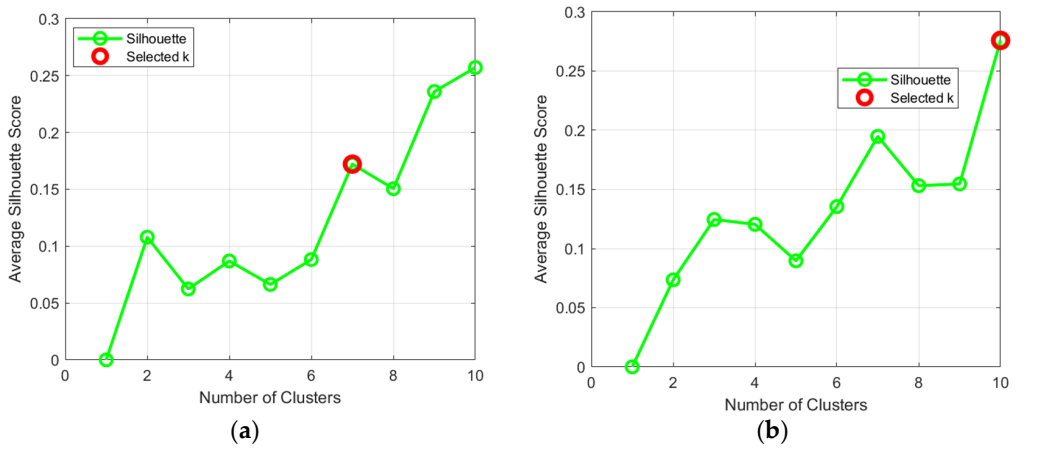


Figure 3. Silhouette analysis comparison for: (**a**) k-means; (**b**) spectral.

The distribution exhibits a reasonable balance with two dominant clusters containing 5 and 7 profiles, respectively, representing 48% of the total prosumer population. Conversely, spectral clustering yielded a 10-cluster solution with a superior silhouette coefficient (0.275), indicating enhanced cluster separation through the algorithm's capacity to capture non-convex patterns in the high-dimensional energy consumption space. However, this improved separation manifests as six singleton clusters, suggesting the presence of unique prosumer behaviours that resist conventional categorization. Table 2 shows a summary of the comparison.

Table 2. Comparison summary for k-means and spectral clustering by metrics.

Optimization Metrics	K-Means	Spectral
Optimal k	7	10
Silhouette coefficient	0.17	0.275
WCSS reduction	47%	50%

6. Analysis of Results

The comparative analysis of clustering methodologies reveals distinct segmentation characteristics impacting community energy management strategies. K-means clustering identified seven clusters with a silhouette coefficient of 0.17, showing a 47% reduction in within-cluster sum of squares. The clusters are reasonably balanced, with two dominant clusters containing 48% of prosumers. Spectral clustering yielded 10 clusters with a superior silhouette coefficient (0.275), capturing non-convex patterns, but resulting in six singleton clusters indicating unique prosumer behaviours.

K-means identifies three prosumer categories: net producers, net consumers, and balanced clusters. EV charging patterns differentiate clusters, with concentrations in 02:00–08:00 and 16:00–23:00 periods. Spectral clustering amplifies distinctions, showing potential for sub-clustering. These findings enable complementary trading pairs for peer-to-peer energy transactions and demand response optimization opportunities through alignment with solar generation peaks. Sub-clustering can refine trading strategies within larger groups.

7. Conclusions

The clustering results reveal the complexity of prosumer behavior in distributed energy systems, highlighting trade-offs between cluster quality and practical implementation. Spectral clustering shows superior mathematical separation over k-means, but its singleton clusters limit applicability for group-based energy trading. In contrast, k-means provides actionable prosumer segments suitable for differentiated tariffs and peer-to-peer trading. Complementary consumption-generation patterns enable temporal arbitrage strategies, with morning surplus in Clusters 1 and 5 potentially offsetting evening deficits in Clusters 6 and 7. The significant variation in cluster sizes and characteristics necessitates adaptive optimization approaches, while singleton clusters indicate outlier prosumers requiring customized strategies. The correlation between EV ownership and net consumption patterns suggests that future EV proliferation will alter community energy balance, requiring dynamic clustering approaches. These findings inform the development of metaheuristic optimization algorithms for energy trading and pricing.

Future work will incorporate a comparison of additional clustering algorithms to inform about the most optimal and dynamic solution, considering homes with solar rooftop PVs and electric vehicle charging stations.

https://doi.org/10.3390/engproc2026140009

Author Contributions: Conceptualization, M.R. and K.A.F.; methodology, M.R., K.A.F. and D.O.; formal analysis, M.R., K.A.F. and D.O.; investigation, M.R.; writing—original draft preparation, M.R.; writing—review and editing, M.R., K.A.F. and D.O.; supervision, K.A.F. and D.O.; funding acquisition, M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work is based on the research supported in part by the National Research Foundation of South Africa (Ref Numbers: YAAP250603319083, CPRR230512105150). The APC was funded by the University of Cape Town and Cape Peninsula University of Technology.

Data Availability Statement: The raw data underlying this article will be made available to readers on request. Sampling data are licensed and can be found on: https://www.pecanstreet.org/dataport/ (accessed on 18 October 2025).

Acknowledgments: Eskom Tertiary Education Support Programme (TESP), South Africa is acknowledged for always supporting research initiatives. During the preparation of this manuscript/study, the authors used Grammarly v1.2.231.1817 for the purposes of correcting language and Meta AI for paragraph content reduction to fit the 8-page limit. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CRE	Community renewable energy
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DER	Decentralized energy resources
DR	Demand response
EV	Electric vehicles
GMM	Gaussian Mixture Model
HC	Hierarchical Clustering
LV	Low voltage
MV	Medium voltage
P2P	Peer-to-peer
PNN	Probabilistic neural networks
PV	Photovoltaic
SOM	Self-organizing maps
WCSS	Within-cluster sum of squares

References

1. Satre-Meloy, A.; Diakonova, M.; Grünewald, P. Cluster analysis and prediction of residential peak demand profiles using occupant activity data. *Appl. Energy* **2020**, *260*, 114246. [CrossRef]
2. Bogensperger, A.; Fabel, Y. A practical approach to cluster validation in the energy sector. *Energy Inform.* **2021**, *4*, 38. [CrossRef]
3. Rajabi, A.; Eskandari, M.; Ghadi, M.J.; Li, L.; Zhang, J.; Siano, P. A comparative study of clustering techniques for electrical load pattern segmentation. *Renew. Sustain. Energy Rev.* **2020**, *120*, 109628. [CrossRef]
4. Michalakopoulos, V.; Sarmas, E.; Papias, I.; Skaloumpakas, P.; Marinakis, V.; Doukas, H. A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs. *Appl. Energy* **2024**, *361*, 122943. [CrossRef]
5. González-Sotres, L.; Frías, P.; Mateo, C. Techno-economic assessment of forecasting and communication on centralized voltage control with high PV penetration. *Electr. Power Syst. Res.* **2017**, *151*, 338–347. [CrossRef]
6. García, L.A.; María, G.; Cardoso, C.; Nowé, A. Two-level clustering methodology for smart metering data*. *Cuad. De Adm.* **2020**, *33*. Available online: https://revistas.javeriana.edu.co/files-articulos/CA/33%20(2020)/20562876001/ (accessed on 3 November 2025). [CrossRef]
7. Užupytė, R.; Krilavičius, T. The Generation of Electricity Load Profiles Using K-Means Clustering Algorithm. *J. Univers. Comput. Sci.* **2018**, *24*, 1306–1329. [CrossRef]

8. Chicco, G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy* **2012**, *42*, 68–80. [CrossRef]
9. Vergados, D.J.; Mamounakis, I.; Makris, P.; Varvarigos, E. Prosumer clustering into virtual microgrids for cost reduction in renewable energy trading markets. *Sustain. Energy Grids Netw.* **2016**, *7*, 90–103. [CrossRef]
10. Han, L.; Morstyn, T.; McCulloch, M.D. Scaling up Cooperative Game Theory-Based Energy Management Using Prosumer Clustering. *IEEE Trans. Smart Grid* **2021**, *12*, 289–300. [CrossRef]
11. Enriquez-Loja, J.; Castillo-Pérez, B.; Serrano-Guerrero, X.; Barragán-Escandón, A. Performance evaluation method for different clustering techniques. *Comput. Electr. Eng.* **2025**, *123*, 110132. [CrossRef]
12. Gbadega, P.A.; Sun, Y.; Balogun, O.A. Optimized energy management in Grid-Connected microgrids leveraging K-means clustering algorithm and Artificial Neural network models. *Energy Convers. Manag.* **2025**, *336*, 119868. [CrossRef]
13. Salehi, N.; Martínez-García, H.; Velasco-Quesada, G. Networked Microgrid Energy Management Based on Supervised and Unsupervised Learning Clustering. *Energies* **2022**, *15*, 4915. [CrossRef]
14. Bellinguer, K.; Girard, R.; Bocquet, A.; Chevalier, A. ELMAS: A one-year dataset of hourly electrical load profiles from 424 French industrial and tertiary sectors. *Sci. Data* **2023**, *10*, 1081. [CrossRef] [PubMed]
15. Jeong, H.C.; Jang, M.; Kim, T.; Joo, S.K. Clustering of load profiles of residential customers using extreme points and demographic characteristics. *Electronics* **2021**, *10*, 290. [CrossRef]
16. Nystrup, P.; Madsen, H.; Blomgren, E.M.V.; de Zotti, G. Clustering commercial and industrial load patterns for long-term energy planning. *Smart Energy* **2021**, *2*, 100010. [CrossRef]
17. Zhan, S.; Liu, Z.; Chong, A.; Yan, D. Building categorization revisited: A clustering-based approach to using smart meter data for building energy benchmarking. *Appl. Energy* **2020**, *269*, 114920. [CrossRef]
18. Zhu, H.; Liao, X.; de Laat, C.; Grosso, P. Evaluation of non-linear power estimation models in a computing cluster. *Sustain. Comput. Inform. Syst.* **2016**, *11*, 26–37. [CrossRef]
19. Pecan Street. Pecan Street Data Port. Available online: https://www.pecanstreet.org/dataport/ (accessed on 18 October 2025).
20. Yin, H.; Aryani, A.; Petrie, S.; Nambissan, A.; Astudillo, A.; Cao, S. A Rapid Review of Clustering Algorithms. *arXiv* **2024**. [CrossRef]
21. Yintong, W.; Wanlong, L.; Rujia, G. An Improved K-means Clustering Algorithm. In *Proceedings of the 2011 IEEE 3rd Int. Conference on Communication Software and Networks*; IEEE: Piscataway, NJ, USA, 2011. [CrossRef]
22. Jin, X.; Han, J. K-Means Clustering. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2010; pp. 563–564. [CrossRef]
23. Shi, J.; Malik, J. Normalized Cuts and Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905. [CrossRef]
24. Ng, A.Y.; Jordan, M.I.; Weiss, Y. On Spectral Clustering: Analysis and an algorithm. In *Proceedings of the 15th International Conference on Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2001; pp. 849–856.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.