

A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs

Vasilis Michalakopoulos^{*}, Elissaios Sarmas, Ioannis Papias, Panagiotis Skaloumpakas, Vangelis Marinakis, Haris Doukas

Decision Support Systems Laboratory, School of Electrical & Computer Engineering, National Technical University of Athens, Greece

ARTICLE INFO

Keywords:

Machine learning
Clustering
Electricity loads
Residential consumers
Demand response

ABSTRACT

Load shapes derived from smart meter data are frequently employed to analyze daily energy consumption patterns, particularly in the context of applications like Demand Response (DR). Nevertheless, one of the most important challenges to this endeavor lies in identifying the most suitable consumer clusters with similar consumption behaviors. In this paper, we present a novel machine learning based framework in order to achieve optimal load profiling through a real case study, utilizing data from almost 5000 households in London. Four widely used clustering algorithms are applied specifically K-means, K-medoids, Hierarchical Agglomerative Clustering and Density-based Spatial Clustering. An empirical analysis as well as multiple evaluation metrics are leveraged to assess those algorithms. Following that, we redefine the problem as a probabilistic classification one, with the classifier emulating the behavior of a clustering algorithm, leveraging Explainable AI (xAI) to enhance the interpretability of our solution. According to the clustering algorithm analysis the optimal number of clusters for this case is seven. Despite that, our methodology shows that two of the clusters, almost 10% of the dataset, exhibit significant internal dissimilarity. As a result, these clusters have been excluded from consideration for DR programs. The scalability and versatility of our solution makes it an ideal choice for power utility companies aiming to segment their users for creating more targeted DR programs.

1. Introduction

As renewable energy sources like solar and wind power become more and more prevalent, the intermittent nature of these sources necessitates a more flexible approach to energy consumption [1]. A direct response to this changing energy landscape is the adoption of Demand Response (DR) mechanisms. DR allows consumers, both residential and industrial, to actively participate in managing the electric grid by adjusting their electricity usage in response to specific signals, typically from utilities or grid operators [2]. In times of peak demand or grid instability, consumers reduce or shift their electricity usage, lessening strain on the grid and preventing potential blackouts [3]. DR not only promotes grid reliability but also contributes to environmental sustainability by synchronizing energy usage with the availability of renewable energy, reducing the need for backup power sources [4]. Additionally, DR creates opportunities for cost savings, as participating consumers can benefit from incentives and lower electricity rates during non-peak hours [5].

Power utilities and grid operators are increasingly using DR programs to help balance supply and demand, ultimately leading to lower

electricity costs in wholesale markets and reduced retail rates [6]. Methods for involving consumers in DR programs include offering time-based pricing options such as time-of-use rates, critical peak pricing, variable peak pricing, real-time pricing, and critical peak rebates [2]. For the effectiveness of the DR programs clusters of consumers are essential [7]. Clusters help streamline the deployment of DR programs by creating homogeneous groups of consumers. By grouping consumers with similar energy consumption patterns, preferences and resources, DR programs can tailor their approaches and incentives, maximizing participation and impact [8]. Clusters also encourage collaboration among prosumers facilitating peer-to-peer energy sharing, promoting local energy resilience, and fostering community engagement in sustainable energy practices [9,10].

Traditionally, power utilities define consumer behaviors using rudimentary methods, relying solely on broad demographic categories and historical consumption data. While these methods provided some insights, they often proved to be insufficiently precise and adaptable as they did not manage to capture the details and complexities that define energy consumption patterns. This lack of precision and agility

^{*} Corresponding author.

E-mail address: vmichalakopoulos@epu.ntua.gr (V. Michalakopoulos).

<https://doi.org/10.1016/j.apenergy.2024.122943>

Received 22 November 2023; Received in revised form 10 February 2024; Accepted 19 February 2024

Available online 6 March 2024

0306-2619/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nomenclature

AI	Artificial Intelligence
CHI	Calinski–Harabasz Index
DBI	Davies–Bouldin Index
DBSCAN	Density-based Spatial Clustering
DL	Deep Learning
DR	Demand Response
FCM	Fuzzy C-Means
HAC	Hierarchical Agglomerative Clustering
ML	Machine Learning
SIL	Silhouette Score
SOM	Self-Organizing Maps
t-SNE	t-distributed Stochastic Neighbor Embedding
ToU	Time-of-Use
xAI	Explainable Artificial Intelligence

hindered the ability of utilities to respond effectively to changing consumer preferences and emerging energy consumption trends and as a result their ability to provide tailored services and effective DR strategies was limited.

The advent of digitization and smart meters has revolutionized consumer behavior analysis. Real-time data from smart meters provide a granular understanding of energy consumption patterns, allowing for more accurate and timely insights [11,12]. For instance, analyzing time of day energy consumption, daily usage pattern stability, and actual volume of energy use can provide valuable insights into household energy consumption habits [13,14]. Moreover, Artificial Intelligence (AI), with its advanced analytics and machine learning capabilities, plays a pivotal role in transforming this data into actionable knowledge [15, 16]. AI algorithms can semi-automatically identify and refine clusters of consumers based on a multitude of factors, such as consumption habits, peak demand periods, and response preferences. Techniques like Hierarchical Agglomerative Clustering (HAC), K-means, Fuzzy C-means (FCM) and Self-Organizing Map (SOM) uncover complex patterns in large datasets, facilitating the creation of more accurate and meaningful consumer clusters [17].

While various clustering techniques and methodologies have been proposed, a significant gap still remains in establishing a systematic approach to identify the ideal number of consumer clusters for the effective design of DR programs. Additionally, there is a need to gain a deeper understanding of the underlying factors contributing to the formation of these segments. This paper seeks to address this critical gap in the existing literature by introducing a systematic, data-driven method for determining the optimal number of clusters when classifying electricity consumers for DR programs, using historical data as the basis. Through this contribution, our research aims to provide valuable practical insights to energy industry stakeholders and utility companies, ultimately facilitating the development of more efficient and tailored DR initiatives, for a greener and more sustainable future in the residential energy domain.

The contributions of our study are summarized as follows:

- Implant thorough description of time-series data through feature engineering, targeted for DR-programs. The pipeline is tested on 4438 houses for more than one year of data, thus showing the generality of our approach in terms of load and time.
- We utilize and compare four different clustering algorithms – namely K-means, K-medoids, Hierarchical Agglomerative (HAC) and Density-based Spatial (DBSCAN) clustering – using various dissimilarity indexes. We assess them using Davies–Bouldin Index (DBI), Calinski–Harabasz Index (CHI) and Silhouette Score (SIL). However, to end up with the prevailing clustering algorithm, in

terms of cluster meaning, we perform an additional similarity matrix and visual evaluation, considering the specific requirements of a practical DR policy, taking into consideration factors such as optimal number of clusters, similarity and dissimilarity.

- We propose a solution that is scalable over time, allowing the integration of a significantly large number of households into the existing classes within mere seconds. This remarkable speed is attributed to the exceptionally low inference response time of the classifier.
- Reinterpretation of the problem as a probabilistic classification one, leveraging Explainable Artificial Intelligence (xAI) to gain deeper insights and build a trusted environment between the power utilities and Machine Learning (ML) solutions. This way, the power utilities can clarify the reason that drives the formation of each user group, giving reliability to the solution, bridging the gap between empirical knowledge and AI.

The rest of the study is organized as follows. In Section 2 a thorough literature review is presented. In Section 3 the developed methodology is described. In Section 4 an experimental application based on a real case study is presented. In Section 5 we delve into an analysis and exploration of the findings, implications and limitations of our study. Finally, concluding remarks are provided in Section 6.

2. Literature review

2.1. Related work

Numerous research papers in the literature have explored the comparative analysis of various clustering techniques within the domain of electricity consumer classification and load pattern segmentation. In Table 1 we present some of the ML methods used for segmenting electricity load consumers.

Chicco et al. conducted a study [31] that comprehensively compares different unsupervised clustering algorithms – HAC, K-means, FCM, Modified Follow the Leader (FDL) and SOM – in terms of their effectiveness in classifying electricity customers based on their consumption patterns. The findings of the study indicate that HAC and FDL methods when run with the average distance linkage criterion offer a more detailed separation of clusters and electricity consumption behavior patterns compared to alternative clustering techniques. In the study by Amin Rajabi et al. [17], among the clustering methods considered, the K-means algorithm emerges as the most commonly utilized, followed by FCM, HAC, and SOM.

The domain of smart meter data has also witnessed the application of artificial intelligence techniques [32]. In the study [33], utility customer segmentation based on smart meter data is empirically examined. The researchers introduce statistical methods that leverage variability measurements to identify consumer segments of varying sizes, which can yield significant outcomes for tailored energy programs. They group the daily profiles of 85 customers using clustering techniques such as K-means, K-medoids, and SOM. This approach sheds light on the fact that predicting individual-level energy consumption is more feasible for customers with stable load profiles and lower variability, as opposed to those with erratic load patterns. Kwac et al. [11] explore the segmentation of household energy consumption patterns using hourly data and propose the use of entropy to capture consumption variability. McLoughlin, et al. [20] present another clustering approach for characterizing domestic electricity load profiles via smart metering data. Their study encompass a comparison of the effectiveness of three commonly used unsupervised clustering methods – K-means, K-medoids, and SOM – based on the DBI. Subsequently, a multi-nominal logistic regression is employed to establish connections between profile classes and various factors such as dwelling type, occupants and appliance characteristics. Czetany et al. [28] conduct a comprehensive analysis of smart meter

Table 1

Literature review on ML methods used for clustering electricity load consumers.

Title	Year	Models	Metrics	Citations
Overview and performance assessment of the clustering methods for electrical load pattern grouping	2012	K-means, FCM, HAC, Follow-The-Leader	Clustering Dispersion Indicator (CDI), DBI, Modified Dunn Index (MDI), Mean Index Adequacy (MIA), Scatter Index (SI), Similarity Matrix Indicator (SMD), Variance Ratio Criterion (VRC), Ratio of Within Cluster sum of squares to between Cluster Variation (WCBCR)	[18]
Deriving the optimal number of clusters in the electricity consumer segmentation procedure	2013	Minimum Variance Method (MVM), FCM	WCBCR, Bayesian Information Criterion (BIC)	[19]
A clustering approach to domestic electricity load profile characterization using smart metering data	2015	K-means, K-medoids, SOM	DBI	[20]
Application of clustering technique to electricity customer classification for load forecasting	2015	K-means, FCM	Proposed clustering validity indicator	[21]
Analyzing the segmentation of energy consumers using mixed fuzzy clustering	2015	Mixed Fuzzy Clustering (MFC)	CHI, DBI, SIL	[22]
A review on clustering of residential electricity customers and its applications	2017	K-means, K-medoids, K-medium, FCM, Adaptive K-means, HAC, SOM, Modified Follow the Leader (FDL), Expectation Maximization (EM)	CDI, DBI, MIA, SMI, Dunn Index (DI), MDI, SIL, WCBCR, SI	[23]
Simulation Study on Clustering Approaches for Short-Term Electricity Forecasting	2018	HAC	CHI, C Index, Duda Index, Ptbiserial Index, DBI, Frey Index, Hartigan Index, Ratkowsky Index, Ball Index, McClain Index, KL Index, SIL, DI, Standard Deviation Index (SDI)	[24]
Comparison of clustering approaches for domestic electricity load profile characterization - Implications for demand side management	2019	K-means	SIL	[25]
A comparative study of clustering techniques for electrical load pattern segmentation	2019	K-means, FCM, HAC, Probabilistic Generative Models (PGM), SOM	Mean Square Error (MSE), SIL, DBI, MIA, WCBCR, DI	[17]
Load Profile Based Electricity Consumer Clustering Using Affinity Propagation	2019	Affinity Propagation (AP)	SIL, CHI, DBI, DI, WCBCR, CDI	[26]
A Novel Clustering Index to Find Optimal Clusters Size With Application to Segmentation of Energy Consumers	2021	K-means, Single Linkage	CHI, DBI, Modified Davies-Bouldin Index (DB**), Gap Statistic (GS), BIC, SIL, S_Dbw	[27]
Development of electricity consumption profiles of residential buildings based on smart meter data clustering	2021	K-means, FCM, HAC	DI, SIL, Elbow method	[28]
An embedded deep-clustering-based load profiling framework	2022	K-means, HAC, Gaussian Mixture Model, Deep Embedding Clustering (DEC), Improved Deep Embedded Clustering (IDEC), Dynamic Autoencoder (DynAE)	SIL, DBI	[29]
K-means clustering of electricity consumers using time-domain features from smart meter data	2023	K-means	SIL, DBI, Elbow method	[30]

data on electricity consumption in Hungarian households. They employ three typical clustering methods, including K-means, FCM and HAC to identify daily and annual energy consumption patterns. Their findings favored K-means over Fuzzy C-means and HAC clustering and four clusters were proposed.

Additionally, Motlagh et al. address the clustering of residential electricity customers through load time series data [13]. They propose a feature-based clustering algorithm that effectively tackles the challenges posed by the high dimensionality of such data. This approach involves transforming load time series into map models, which can be easily clustered, thereby mitigating the limitations associated with extreme dimensionality. By clustering customers based on their load patterns, valuable insights into customer preferences, usage patterns, and energy demand characteristics are gleaned. This facilitates the design of tailored electricity sales packages that align with the consumption patterns of each cluster, ultimately leading to enhanced customer satisfaction. Devijver et al. [34] explore the clustering of residential electricity consumers utilizing high-dimensional regression

mixture models. This involves the calculation of clusters followed by the fitting of predictive models for each cluster.

Furthermore, several novel Deep Learning (DL) clustering algorithms customized for electricity customer classification and load pattern segmentation have been proposed. For example, [35] introduces a clustering algorithm that combines the hopfield neural network and K-means algorithm to enhance electricity customer segmentation accuracy. The hopfield artificial neural network is utilized to eliminate the random generation of initial cluster centers. In another study, the Iterative Refinement Clustering (IRC) method merges desirable properties of hierarchical clustering and the Follow-the-Leader approach to achieve effective clustering outcomes [36]. This method operates with a predetermined number of clusters and incorporates a specific mechanism to prevent the creation of empty clusters. Zarabie et al. [26] assert the superior grouping capabilities of the Affinity Propagation (AP) algorithm compared to other methods such as K-means, K-medoids, and spectral clustering in terms of grouping residential load profiles. In their study [29], Eskandarnia et al. explore DL clustering algorithms like

Deep Embedding Clustering (DEC), Improved Deep Embedded Clustering (IDEC), and Dynamic Autoencoder (DynAE), based on autoencoders on two real-world datasets namely the 5567 London Households dataset and the 5000 Irish Households dataset. Comparatively, traditional machine learning methods – K-means, HAC, Gaussian Mixture Model – were employed to benchmark the autoencoder-based algorithms. The DynAE effectively grouped consumers into four clusters, providing unique load profiles for each group. These innovative approaches aim to improve the accuracy and efficiency of clustering techniques in the electricity domain.

At last, studies have explored the optimal number of clusters in the electricity consumer segmentation process. Nameer Al Khafaf et al. [27] propose a clustering index derived from eigenvalue analysis on the correlation matrix of smart meter data to determine the optimal cluster size. To control the number of features, they also incorporate a genetic algorithm-based feature selection process. Similarly, [19] suggests methods for optimizing cluster numbers by considering factors such as geographical characteristics, voltage level, and demographic features. Wang et al. [37] introduce an algorithm based on DBSCAN and K-means for consumer segmentation as a potential application in the Yunnan Electricity Market. The optimality of the clustering result is evaluated by the SIL, CHI and DBI indexes. In parallel Zhang et al. [38] propose a method to analyze the power consumer behavior based on the K-means and DBSCAN. In another study [30], researchers employed unsupervised classification techniques on real data from a sample of 5567 London Households, to group consumers based on their similar electricity consumption patterns. Their approach involved utilizing the K-means clustering algorithm to identify consumption patterns among households. The optimal cluster number was determined through the elbow method, with cluster validity assessed using the SIL coefficient and DBI. Their findings revealed that the average electricity usage for all users is depicted across three clusters.

After conducting a review of the existing literature we identified significant gaps in the interpretation of formed clusters and the scalability of the methods employed. Many studies present segmentation methods that appear effective on a small scale but lack consideration for their applicability to larger populations. As power utilities increasingly seek to engage a broader range of consumers in DR initiatives, it is crucial to develop scalable segmentation methods that can support the complexity and diversity of the utility's consumer base. Additionally, while studies may successfully identify distinct consumer groups based on various characteristics such as consumption patterns and demographics, they often fall short in explaining the reason why these clusters were formed and how they can translate into actionable strategies for utilities. Furthermore, the presented solution handles a large dataset producing meaningful results, whilst deciding the best number of clusters for demand response programs in a unique way to other studies.

In this paper, we address the identified knowledge gaps in the literature by proposing an innovative solution that compares K-means, K-medoids, HAC and DBSCAN algorithms using DBI, CHI and SIL indexes. In order to determine the most suitable algorithm based on cluster meaning we conduct a similarity matrix and visual assessment with attention to DR needs. To provide a scalable solution we utilize a classifier for rapid household integration into clusters and xAI for detection of the reasons behind each user group's formation.

3. Methodology

The proposed methodology is presented in Fig. 1. First, a detailed description of the first phase of the pipeline that extracts the features from the given data is issued. Following, clustering algorithms and the appropriate evaluation metrics are utilized to obtain the optimal number of clusters. In the next subsection, similarity matrices between the different algorithms are used to obtain the nominal clusters produced. On top of those clusters, a probabilistic classification function and xAI part are introduced as the final step of our proposed approach.

More specifically, the first phase of the process begins with load profiling of different residential households, capturing essential energy usage patterns and creating new features to enable data-driven decision-making. The obtained load profiles are then subjected to various clustering algorithms, utilizing evaluation metrics to determine the optimal number of clusters for accurate representation. To ensure the robustness of the results, similarity matrices are employed to assess the consensus among those clustering algorithms and produce the nominal clusters. Subsequently, the problem is reinterpreted as a probabilistic classification challenge and xAI techniques are used for the creation of additional specific clusters, failed to be identified in the previous step, facilitating deeper insights into energy consumption behavior. Through this seamless combination, stakeholders can gain comprehensive knowledge of household energy usage, identify energy saving opportunities, and implement effective DR programs to achieve environmental and economic benefits. Furthermore, the transparency facilitated by advanced xAI can bolster the confidence of power utilities and offer them a clear view of the key factors influencing the categorization of user groups, thus promoting well-informed decision-making processes.

3.1. Load pattern extraction

In this phase, energy consumption data from different residential buildings are initially collected and analyzed to create load profiles. Load profiling refers to the careful analysis of how energy is consumed or used over a specific period. It involves a detailed examination of patterns and behaviors in energy consumption during that defined timeframe. This examination involves identifying peak consumption periods, overall statistics of the consumption, and providing a detailed characterization of the overall usage patterns.

Through an extensive literature review, three distinct groups of attributes have been identified as the most important inputs for the subsequent clustering. The first group encompasses the raw time-series data, representing the average energy consumption for each examined house in 30-minute intervals throughout the day. The second group comprises features to pinpoint peak energy consumption during different periods throughout the day, namely the **peak early morning**, **peak morning**, **peak noon**, **peak evening**, **peak night**, and **peak late night**. More specifically for the mentioned time periods:

- **peak early morning**: Represents the time frame from 05 : 00 AM to 10 : 00 AM, signifying the early morning peak in energy usage.
- **peak morning**: Encompasses the hours from 10 : 00 AM to 02 : 00 PM, identifying the morning peak in energy consumption.
- **peak noon**: Corresponds to the period between 02 : 00 PM and 05 : 00 PM, highlighting the midday peak in energy usage.
- **peak evening**: Covers the time span from 05 : 00 PM to 09 : 00 PM, indicating the evening peak in energy consumption.
- **peak night**: Extends from 09 : 00 PM to 12 : 00 AM, representing the nighttime peak in energy usage from 09 : 00 PM to midnight.
- **peak late night**: Spans from 12 : 00 AM to 05 : 00 AM, indicating the late-night peak in energy consumption, covering the early morning hours.

The third group encompasses statistical descriptors, such as mean of the time-series, std (standard deviation), min, max, 25th percentile, 50th percentile, and 75th percentile. These statistical measures offer a comprehensive view of data variability, and distribution characteristics, further enhancing our understanding of energy consumption patterns in the specific dataset. Leveraging these three distinct groups of attributes yields valuable insights into the intricacies of energy consumption behaviors, forming the cornerstone for subsequent clustering and classification algorithms.

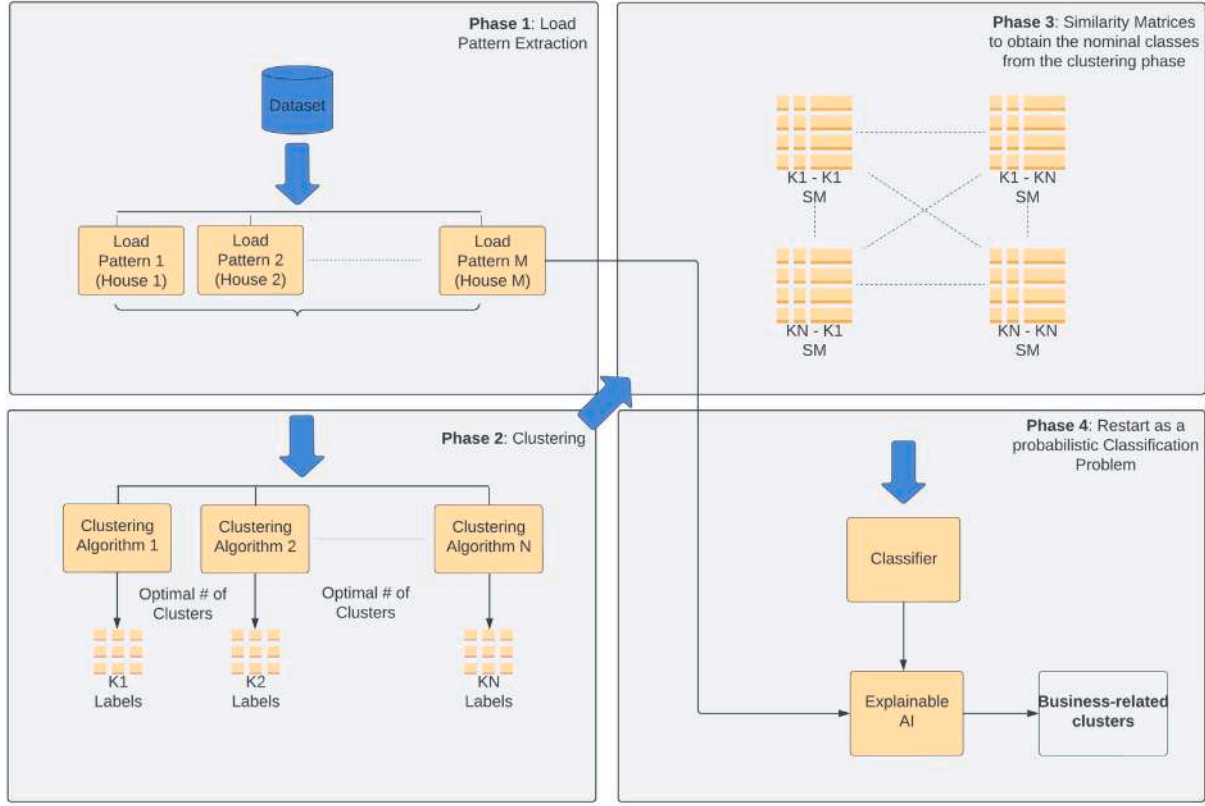


Fig. 1. Proposed methodology.

3.2. Clustering algorithms

Once the load profiling process has been completed, the next step involves applying various clustering algorithms, as an unsupervised learning technique, to group together similar energy consumption patterns within the data. Different clustering techniques, such as K-means, K-medoids, HAC and DBSCAN clustering, are employed to partition the data into distinct clusters. Each of these algorithms offers a unique approach to identifying patterns and structures within the energy consumption data, allowing for a comprehensive exploration of potential groupings. By applying multiple algorithms to the same data, it is possible to assess which one yields the best results in terms of cluster quality, inner-similarity and dissimilarity. The drawbacks and advantages of each algorithm, as applied in the proposed methodology, are described in the following paragraphs.

3.2.1. K-means

The K-means algorithm [39] is one of the most widely used unsupervised learning technique in machine learning and it is commonly employed for clustering analysis, where a given dataset is divided into k clusters. The minimization of an objective function is the main goal of the K-means algorithm. This function represents the sum of the squared distances between each data point and its assigned cluster center.

Mathematically, the aforementioned function J is defined in (1):

$$J = \sum_{i=1}^k \sum_{j=1}^n \|x_j - c_i\|^2 \quad (1)$$

Where J represents the objective function, k is the number of clusters, c_i is the centroid of cluster i , x_j is a data point, and $\|x_j - c_i\|$ denotes the Euclidean distance between data point x_j and cluster centroid c_i .

The K-means algorithm operates through an iterative process that involves the following steps:

1. Initialization: Randomly select k initial centroids or use a predefined initialization method.
2. Assignment: Assign each data point to the nearest centroid, forming k clusters.
3. Update: Recalculate the centroids of each cluster as the mean of all data points in that cluster.
4. Convergence: Repeat the assignment and update steps until convergence is achieved, i.e., when the centroids no longer change significantly or a predefined convergence criterion is met.

The K-means algorithm is known for its simplicity, ease of implementation, and computational efficiency, which makes it suitable for analyzing large datasets. Over the years, various modifications and improvements have been proposed, including the use of different distance metrics, initialization methods, and convergence criteria, to enhance its performance in various applications.

3.2.2. K-medoids

The K-medoids algorithm [40], is a variation of the K-means algorithm. It is widely used for clustering analysis and has the advantage of being robust to outliers. K-medoids, unlike K-means, uses data points themselves (medoids) as cluster representatives. The medoid of a cluster is defined as the specific piece of data that exhibits the lowest average dissimilarity when compared to all other points within the same cluster.

The main objective of the K-medoids algorithm is to partition the dataset into k clusters in a way that minimizes the total dissimilarity or cost. This cost is typically calculated as the sum of distances (using the Euclidean distance metric) between each data point and its assigned medoid.

Mathematically, the objective function J for K-medoids can be represented in Eq. (2):

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} d(x_j, m_i) \quad (2)$$

Here, J represents the objective function, k is the number of clusters, C_i is cluster i , x_j is a data point in cluster C_i , m_i is the medoid of cluster C_i , and $d(x_j, m_i)$ is the distance between data point x_j and medoid m_i .

The K-medoids algorithm proceeds as follows:

1. Initialization: Select k initial medoids, either randomly or through a predefined method.
2. Assignment: Assign each data point to the nearest medoid, forming k clusters.
3. Update: For each cluster, choose the data point as the new medoid that minimizes the total cost (dissimilarity) within that cluster.
4. Convergence: Repeat the assignment and update steps until convergence is achieved. Convergence occurs when the assignment of data points to clusters and the selection of medoids no longer change significantly.

The K-medoids algorithm is particularly useful when dealing with datasets that contain outliers, as it is less sensitive to outliers than K-means, due to its nature. It also offers flexibility in terms of the choice of distance metric, making it suitable for various types of data and applications. However, besides the advantages it has over K-means, it is more computationally expensive than the K-means algorithm, especially for large datasets.

3.2.3. Hierarchical Agglomerative Clustering (HAC)

Agglomerative Clustering is a hierarchical clustering technique that initiates by assigning each data point to its own cluster. It is also known as the bottom-up approach or Hierarchical Agglomerative Clustering (HAC). The assumption is that data points that are close to each other are more similar or related than data points that are farther apart. A dendrogram, a tree-like figure, is created by iteratively merging or splitting clusters based on a measure of similarity or distance between data points. HAC is quite flexible as it allows different linkage criteria to calculate the distance between pieces of data, and clusters by extension. Namely:

- Single linkage: Measuring the distance between the nearest pair of points from different clusters.
- Complete linkage: Evaluating the distance between the farthest pair of points from different clusters.
- Average linkage: Calculating the average distance between all pairs of points from different clusters.
- Ward linkage : Minimizes the variance of the clusters being merged.

In mathematical terms, the agglomerative clustering algorithm can be described as shown in (3):

$$d_{ij} = \text{linkage}(C_i, C_j) \quad (3)$$

where d_{ij} signifies the distance between clusters C_i and C_j , and *linkage* is the chosen linkage criterion. HAC has several advantages, including its capability to handle non-linearly separable data and its interpretability through the dendrogram. Nevertheless, in the downsides section, it tends to be computationally intensive for sizable datasets and can exhibit sensitivity to noisy data.

3.2.4. Density-based spatial clustering (DBSCAN)

DBSCAN [41] is a density-based clustering algorithm that is particularly effective for identifying clusters with varying shapes and handling noisy data. It operates by defining clusters as dense regions of data separated by sparser areas, unlike the previous techniques.

The main objective of the DBSCAN algorithm is to partition the dataset into clusters such that each cluster is a dense region separated from other clusters by areas of lower point density. DBSCAN has several advantages, including its ability to discover clusters of varying shapes

and sizes and its robustness to noise. It does not require specifying the number of clusters in advance, making it suitable for datasets with an unknown or variable number of clusters. Moreover, the choice of ϵ (epsilon) and *MinPts* parameters can significantly impact the clustering results. Tuning these parameters is often necessary to achieve the desired clustering outcome.

3.3. Evaluation metrics

To determine the optimal number of clusters for the data, the application of evaluation metrics becomes indispensable. Widely used metrics in the scientific literature, such as the SIL, DBI and the CHI were utilized for evaluating the quality of the clustering results. They assist in selecting the most suitable number of clusters that effectively capture and represent the inherent patterns in the data. By leveraging these evaluation metrics, researchers and practitioners can make informed decisions about the clustering outcomes and refine their analyses accordingly, enhancing the overall quality of the load profiling process.

3.3.1. Silhouette score (SIL)

SIL [42] is probably the most popular measure to evaluate cluster validity. Considering the mean intra-cluster distance, denoted as a , and the mean nearest-cluster distance, denoted as b , for each sample. SIL for an individual sample is given by the formula:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (4)$$

where b_i is the distance between the sample and the nearest cluster that the sample does not belong to. It is crucial to note that the SIL Coefficient is only defined when the number of labels satisfies the condition $2 \leq n_{\text{labels}} \leq n_{\text{samples}} - 1$ and that SIL can range from -1 to 1 . We execute the clustering algorithms across a spectrum of cluster numbers, ranging from 2 to 32, and compute SIL for each outcome. The optimal number of clusters is determined by selecting the clustering configuration that yields the highest SIL.

3.3.2. Davies–Bouldin Index (DBI)

The Davies–Bouldin Index (DBI) [43] is employed as a vital evaluation metric to determine the optimal number of clusters, denoted by K . The index is defined as follows:

$$DB = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \left(\frac{S_i + S_j}{d_{ij}} \right) \quad (5)$$

where:

- S_i is the average distance between each data point within cluster i and its centroid.
- d_{ij} is the distance between the centroids of clusters i and j .

Again, we perform the clustering algorithms with a range of cluster numbers, from 2 to 32, and calculate the DBI for each result. The optimal number of clusters is selected based on the clustering configuration that yields the lowest DBI. A lower index indicates better-defined clusters with reduced overlap, making it an appropriate criterion for identifying the optimal number of clusters.

3.3.3. Calinski–Harabasz Index (CHI)

The Calinski–Harabasz Index [44] which was proposed in 1974 by Calinski T and Harabasz J, (also known as the Variance Ratio Criterion) is used as a significant evaluation metric to determine the optimal number of clusters, denoted by K , in the clustering process. It is defined in (6)

$$CH = \frac{B(K)}{W(K)} \times \frac{N - K}{K - 1} \quad (6)$$

where:

- $B(K)$ is the between-cluster sum of squares, representing the dispersion between the cluster centroids.
- $W(K)$ is the within-cluster sum of squares, measuring the compactness of data points within clusters.
- N is the total number of data points.

Similar to the aforementioned metrics, we apply the clustering algorithms with various cluster numbers, ranging from 2 to 32, and calculate the CHI for each result. The optimal number of clusters is selected based on the clustering configuration that achieves the highest CHI. A higher index value signifies well-separated and compact clusters, making it an appropriate criterion for identifying the optimal number of clusters.

3.4. Similarity matrices

After obtaining the clustered energy consumption patterns, various algorithms are used for the clustering process. It is essential to evaluate the consistency and agreement between these different algorithms. To achieve this, similarity matrices are constructed to compare the labels assigned to data points in each cluster. A high similarity score indicates a strong consensus among the algorithms, ensuring potent results in the clustering outcomes and obtaining the nominal groups. On the other hand, having low similarity in cluster label indicates that this cluster might not be an accurate fit for the specific use case, despite the algorithms previously selecting it.

Adding to this, we employ t-SNE [45] (t-distributed Stochastic Neighbor Embedding) as a crucial tool to visualize and interpret our clustering results. t-SNE is a dimensionality reduction technique that excels in revealing the underlying structure of high-dimensional data by projecting it into a lower-dimensional space while preserving pairwise similarities. This reduction not only simplifies the representation but also enhances our ability to discern distinct clusters and patterns within the data. By utilizing t-SNE, we can transform complex multi-dimensional data into a two-dimensional space, like our case, where the clusters become visually apparent, enabling a more intuitive and insightful understanding of our clustering outcomes. This visualization not only assists in validating the quality of our clustering algorithm but also in generating valuable insights that enrich the broader context of our research, alongside the examination of similarity matrices.

3.5. Classification & xAI

In this phase, additional clusters are created by incorporating business-related features and domain-specific information. After thorough experimentation with popular algorithms such as XGBoost [46] and LightGBM [47], we opted for the CatBoost Classifier [48]. This choice is not inherited only by the goodness-of-fit on the clustered data but also unlike its counterparts, CatBoost stands out for its versatility in handling big data and categorical features, as highlighted in a comprehensive review [49] and a similar classification task [50].

CatBoost, proposed by a team of engineers from Yandex in 2017 is an efficient classifier algorithm that uses gradient boosting on decision trees and handles categorical features in the data. Furthermore, categorical features are not handled during processing but are instead addressed during the training phase. Rather than employing binary substitution for categorical data, it utilizes random permutations and calculates a mean label value. To illustrate, similar class values are positioned ahead of the known ones in the permutation.

Additionally, while clustering algorithms are effective in grouping data points into clusters based on their similarities, they often struggle with generalization when applied to new, unseen data. This is where a classifier comes in handy. By training a classifier to perform the same function as the clustering algorithm, we enable it to make predictions on new data points and assign them to the appropriate clusters, thus extending the clustering model's ability to generalize. Needed to be

said, the data points are of an unlabeled nature before the clustering algorithms take place and segment them into categories. As seen in 1, the phase four of our pipeline redefines the problem as a classification one, using the labeled data produced by the clustering algorithms and their filtering during phase 3. This combined approach not only enhances the accuracy and robustness of the clustering results but also provides a more versatile and adaptable framework for handling diverse data, making it a valuable tool in the case of electrical energy behavior analysis.

In traditional clustering methods, you might obtain clusters, but they might seem arbitrary or lack context. You have clusters, but you are not entirely sure why certain data points ended up together or how these clusters relate to your specific business goals. Due to this fact, xAI methods are employed to interpret the classifiers results and provide insights into the probabilities of each cluster. That way we can provide new clusters with actual business value and create them according to the objectives of the whole clustering approach.

4. Case study

4.1. Data set

Open energy consumption data were used for the examined study that originated from 5567 households in the city of London that span from November 2011 to February 2014.¹ The dataset captures energy readings at half-hour intervals and includes the energy consumption (kWh), the respective timestamp and a household identifier. Around 1100 of the households were exposed to dynamic Time of Use (ToU) energy prices throughout 2013. These customers received advanced notifications of different price signals through Smart Meter IHD or mobile phone messages. This setup aimed to simulate potential future scenarios for effectively managing high renewable energy generation and leveraging high prices to ease the strain on local distribution grids during critical periods. In contrast, about 4500 customers who were not enrolled in ToU tariff schemes were considered more appropriate for examination due to their unfiltered data, which provides a better representation of real-world circumstances. Fig. 2 presents the average energy consumption of the examined houses throughout the entire duration of data collection. As it can be seen, different groups of houses can be distinguished in terms of the volume of energy consumption. In addition, there are notable energy consumption peaks in the day, with most occurring early in the morning and in the evening, which can be attributed to the specific consumption pattern of the users for each household.

Real-world data often exhibit imperfections, such as incompleteness, inconsistency, inaccuracy, as well as the presence of errors, outliers, or missing values. Before delving into the data analysis, the data cleaning process was conducted to enhance data quality and uncover meaningful relationships within the dataset. Initially, data points containing missing or erroneous values were removed to ensure data integrity, resulting to the final number of households reviewed in this study, 4438. Subsequently, a normalization procedure was employed, utilizing the Min-max normalization method to scale all values within the range of 0.0 to 1.0. More specifically, for every feature, the minimum value of that feature gets transformed into a 0, the maximum value gets transformed into a 1, and every other value gets transformed into a decimal between 0 and 1. This normalization step plays a pivotal role in mitigating the potential impact of varying data magnitudes on subsequent clustering analyses, thereby averting associated biases. The applied data preparation procedure aims to enhance the robustness of the subsequent clustering analyses by normalizing data scales, while also facilitating the application of distance-based metrics in data exploration.

¹ <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>

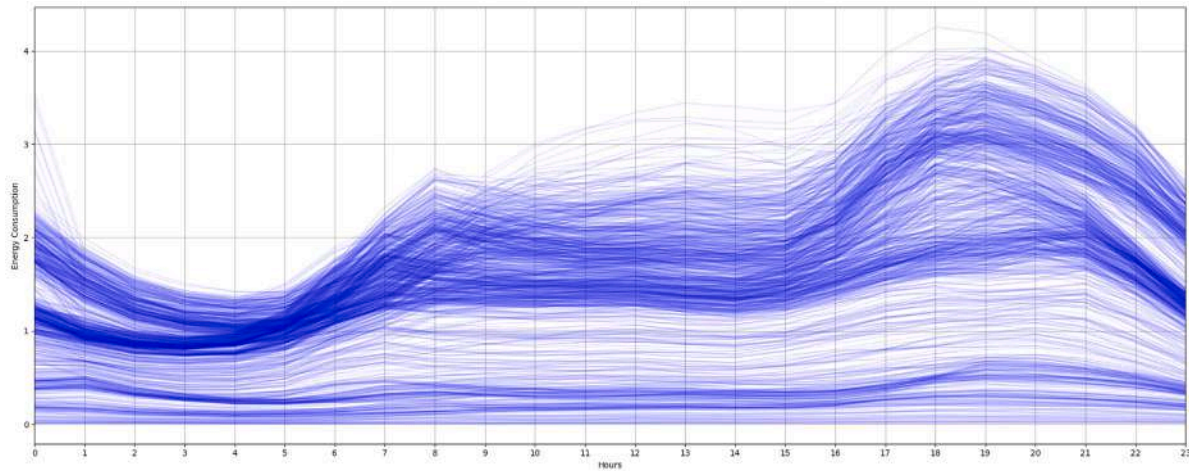


Fig. 2. Average daily energy consumption for all households.

Table 2

Optimal number of clusters for each algorithm based on three evaluation indexes.

Clustering algorithm	SIL	DBI	CHI
K-means	7	7	7
HAC	7	7	8
K-medoids	2	2	7

4.2. Experimental design

After the cleaning process of the utilized data, the proposed pipeline is applied. The first step involves a careful analysis of the energy consumption patterns in the examined households. This analytical phase is pivotal in gaining insights into the temporal dynamics of energy usage for each consumer throughout the years.

Moving into the second part of the proposed framework, the four clustering algorithms, mentioned in 3.2, are applied to segment the data based on the features of each load profile. In the context of clustering analysis, determining the optimal number of clusters is a challenging task, as it typically cannot be precisely known in advance. Consequently, the various clustering algorithms are subjected to testing over a predefined range, spanning from 2 to 30 clusters. This broad range of cluster numbers is explored systematically to identify the most suitable or “best” number of clusters based on three evaluation metrics, including the SIL, DBI and CHI, as mentioned in 3.3.

In Fig. 3, we can see the difference in SIL scores for all the aforementioned algorithms, except DBSCAN, for every number of clusters tested. As can be seen from the diagram, the optimal number of clusters based on this metric is seven. The only exception can be found in 3(b), where the number seven comes second in rank. Following this, the SIL scores for the K-medoids algorithm are too low compared to the other two algorithms and thus we do not consider this diagram quite representative to our case.

As previously mentioned, DBSCAN is a density-based clustering algorithm that does not necessitate users to pre-define the number of clusters as an input parameter. Instead, DBSCAN dynamically determines the number of clusters based on the epsilon (ϵ) value and other criteria. Remarkably, during empirical testing with various epsilon values, it consistently identifies seven as the optimal number of clusters. Consequently, we select the cluster labels associated with the highest evaluation metrics across a range of epsilon values.

Adding to the previous, as we can see in Table 2 most of the algorithms and their corresponding indexes agree that the number seven is the optimal one, in terms of load profile clustering, with the given features. Another useful visual that can be used to reduce the dimensions of the problem and give further insights into the clusters

similarity and dissimilarity employing t-SNE, can be seen in Fig. 4. As we can witness, the clusters are quite separated but improvements can be made. This is due to the fact, that some points that belong to different clusters are quite close in this diagram and can probably be a part of some other cluster or create a new, not already formulated one. Nevertheless, we anticipate confirmation from the similarity matrices to discern that the blue and red clusters are indeed the primary ones identified by the algorithms. Moreover, it is noticeable from this visualization that the cluster points highlighted in light blue exhibit the highest degree of dissimilarity. This observation suggests that there is room for enhancement in terms of improving cluster similarity within this specific cluster.

In Fig. 5, we present a visual representation of the clustering results obtained from our data analysis, on top of the time-series data. Each line showcases the division of data points into seven distinct clusters, which were generated using the K-means algorithm. Each cluster is denoted by a unique color for visual clarity. It is evident that the data points within each cluster share common characteristics and exhibit a certain level of intra-cluster similarity, depending mostly on the daytime peaks of the time-series. The distribution of data points across these clusters provides initial insights into the underlying structure of our dataset and the distinct cluster patterns it encompasses.

The decision to combine all clusters into a single figure serves as a powerful visualization strategy to showcase the antagonism (dissimilarity) between clusters. This choice is influenced by several key factors, with a primary emphasis on the incorporation of density information. Density serves as a fundamental metric, offering valuable insights into the concentration of data points within each cluster. The figure, by showcasing the density profiles of distinct clusters, aids in the identification of clusters exhibiting different levels of dispersion.

Adding to the previous, Fig. 6 presents a comprehensive visualization of all clusters, each displayed in separate sub-figures, accompanied by essential statistical measures such as density, mean, and median. This integrated representation offers a holistic view of the clustering results, enabling a better understanding of each cluster. Moreover, the inclusion of mean and median values for each cluster's features further enhances the interpretability of the clusters. The mean provides a measure of central tendency, indicating the average behavior of data points within a cluster, while the median offers insights into the distribution's central point, the starting point for every center-based clustering algorithm.

Moving on to the similarity matrices between the algorithms, Table 3 illustrates the comparative performance of each clustering algorithm. Each row and column represent a cluster, and the values within the matrix reveal the degree of similarity between these clusters for the corresponding algorithms in comparison. Notably, as can be seen in

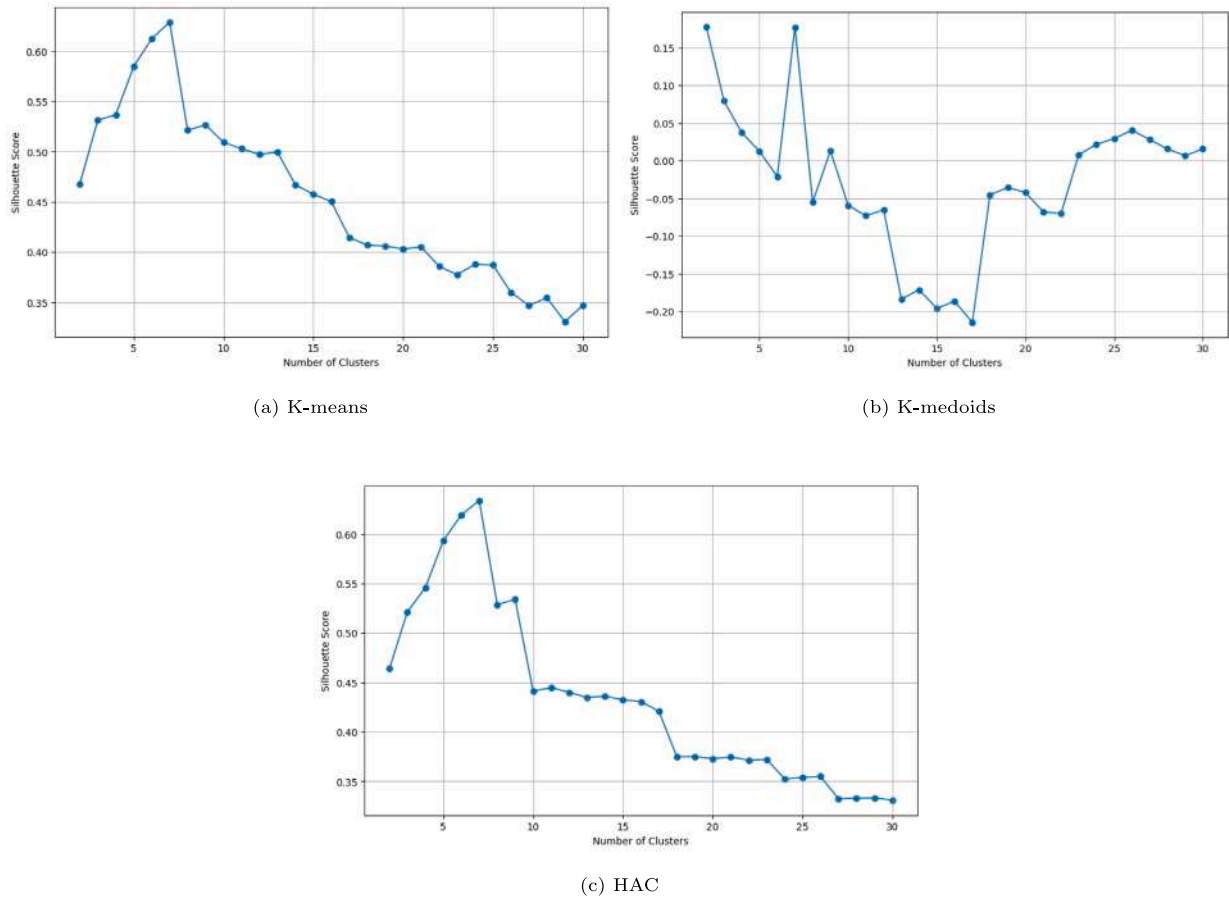


Fig. 3. SIL score for every number of clusters and algorithm.

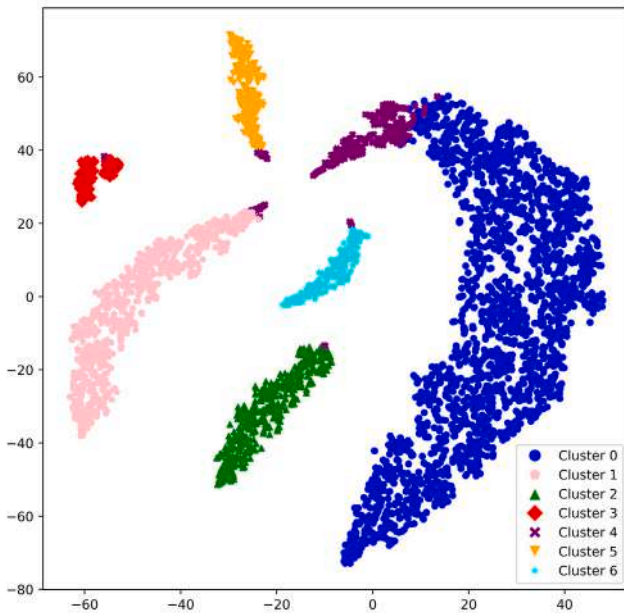


Fig. 4. 2D representation of the clusters for K-means labels.

Table 3a, clusters 0 and 1 in the K-means vs. HAC comparison exhibit a high degree of similarity, suggesting consistent and robust performance across these algorithms for these specific clusters. However, Cluster 4

in the same comparison shows a significant dissimilarity, indicating challenges in achieving agreement.

Similarly, the comparison between K-means and K-medoids reveals strong similarities in clusters 1 and 4, but also highlights variations in other clusters, as can be seen in Table 3b. This divergence underscores the inherent differences in these two methods, with K-means prioritizing cluster centroids and K-medoids focusing on medians, leading to variations in cluster assignments. As expected from the low validation indexes produced by K-medoids, it is obvious that the latter is not suitable for our case and it will no longer be under consideration.

Table 3c depicts the assessment of HAC against DBSCAN, where distinct patterns of cluster formation emerge. The two algorithms agree almost in every data point for clusters 0,1 and 2, showing again that those are the nominal classes as previously shown in the other tables. To the contrary, classes 3 of HAC and 6 of DBSCAN exhibit significant disagreement on data points between the algorithms. This discrepancy suggests that there is indeed room for improvement.

In our comparative analysis of HAC, K-means, and DBSCAN, using the provided similarity matrices, intriguing insights into cluster performance have emerged. Above all, while HAC and DBSCAN exhibit alignment in certain clusters, such as cluster 2, 1 and 0, where they agree strongly, K-means often deviates from this consensus, particularly in cluster 4. This divergence highlights opportunities for improvement in the cluster results. For instance, clusters where K-means does not concur with other algorithms may benefit from further refinement. Conversely, the regions where all three algorithms agree, such as cluster 0,1,2,3 and 6 of K-means, provide a stable foundation for clustering analysis. On the other hand, clusters 4 and 5 indicate signs that further improvements are needed.

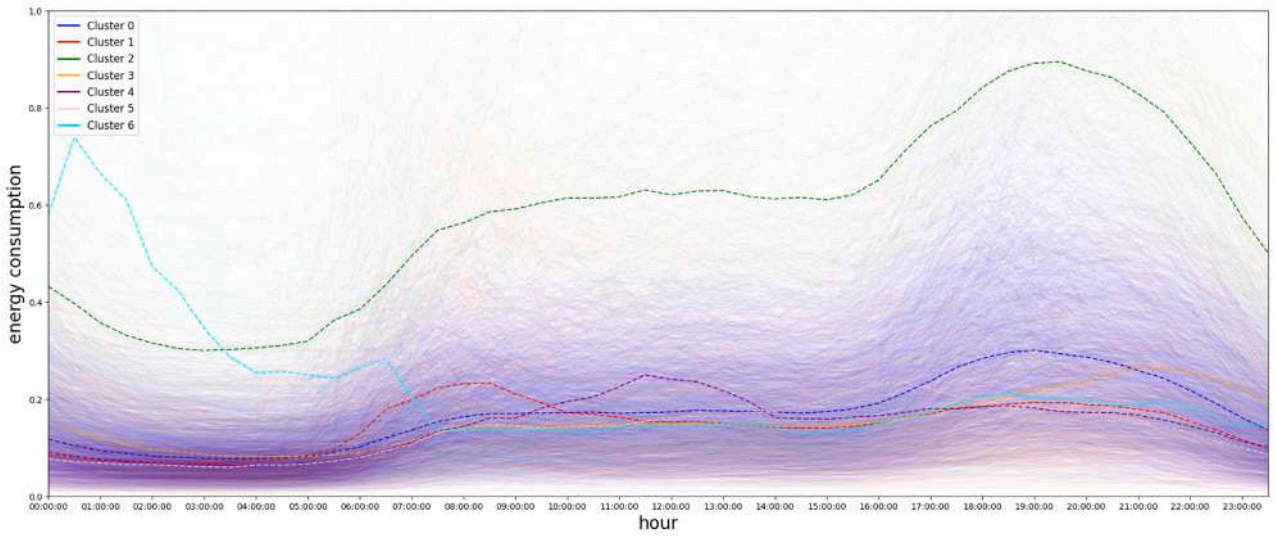


Fig. 5. Median values for each cluster during the day.

Table 3
Comparison of Similarity Matrices.
(a) K-means vs. HAC

	0	1	2	3	4	5	6
0	2442	0	0	0	0	0	0
1	0	0	485	0	0	0	0
2	0	0	0	0	417	0	0
3	0	0	0	0	0	239	0
4	111	9	6	123	0	0	0
5	0	399	0	0	0	0	0
6	0	0	0	1	0	0	206

(c) HAC vs. DBSCAN

	0	1	2	3	4	5	6
0	2553	0	0	0	0	0	0
1	0	126	0	0	278	0	4
2	0	0	0	0	0	0	491
3	81	0	0	12	3	2	26
4	0	0	416	0	0	0	1
5	0	0	0	238	0	0	1
6	0	0	0	0	0	206	0

(b) K-means vs. K-medoids

	0	1	2	3	4	5	6
0	557	223	509	660	0	242	251
1	0	0	0	0	475	10	0
2	0	0	0	0	414	3	0
3	0	0	0	0	235	4	0
4	0	0	0	0	4	245	0
5	0	0	0	0	396	3	0
6	0	0	0	0	205	2	0

(d) K-means vs. DBSCAN

	0	1	2	3	4	5	6
0	2442	0	0	0	0	0	0
1	0	0	0	0	0	0	485
2	0	0	416	0	0	0	1
3	0	0	0	238	0	0	1
4	192	4	0	12	7	1	33
5	0	122	0	0	274	0	3
6	0	0	0	0	0	207	0

4.3. Results

In the previous chapter, after comparing the clustering algorithms, data points have been individually labeled based on the outcomes obtained from each algorithm. Now that we have reached the final phase of our proposed methodology, we introduce the integration of a CatBoost classifier as a pivotal component of our analytical pipeline. This classifier is strategically employed to complement the existing clusters generated by our clustering algorithms. The classifier always learns from the clustering algorithm perfectly and thus it behaves exactly like the corresponding unsupervised learning technique.

The labeled data sets are split into train and test parts, with an 80–20 split ratio. Then, test parts are hidden from the models during the training phase. In order to obtain class probabilities and enable probabilistic class interpretation, the CatBoost's output is subject to a probability calibration process. In addition to making class predictions, this calibration procedure calculates the likelihood associated with each respective label through the classification process. These probabilities, computed for each data point, serve to provide a measure of confidence in the prediction outcomes. This application is made to add a new function to the clustering algorithms to return the probability value for each data point that is decided to belong to a certain cluster.

Now for the baseline of our approach we use the labels produced by the K-means algorithm as the training and test set, in order to see

Table 4
CatBoost Algorithm Performance Metrics per class.

Classes	Precision	Recall	F1 Score
Class 0	0.99	1.00	1.00
Class 1	1.00	0.99	0.99
Class 2	1.00	1.00	1.00
Class 3	1.00	1.00	1.00
Class 4	0.95	0.91	0.93
Class 5	1.00	1.00	1.00
Class 6	1.00	1.00	1.00

the accuracy of the classifier. From Table 4, we can see that CatBoost fits almost perfectly to the given data for every cluster. In this case, while it is evident that CatBoost's near-perfect fit to the data may be indicative of over fitting, it aligns well with our objectives, as we need the classifier to be mimicking the clustering algorithm. The only difference is that the user groups, will be mentioned as classes instead of clusters from now on, due to the change in problem formulation.

So from now on, we have a classifier that behaves almost identically to a clustering algorithm. Utilizing widely recognized metrics like Precision, Recall, and F1 Score, we can discern that the CatBoost algorithm exhibits remarkable performance across various classes. However, it is noteworthy that in most classes, including Class 0, Class 1, Class 2, Class 3, Class 5, and Class 6, the algorithm achieves exceptionally

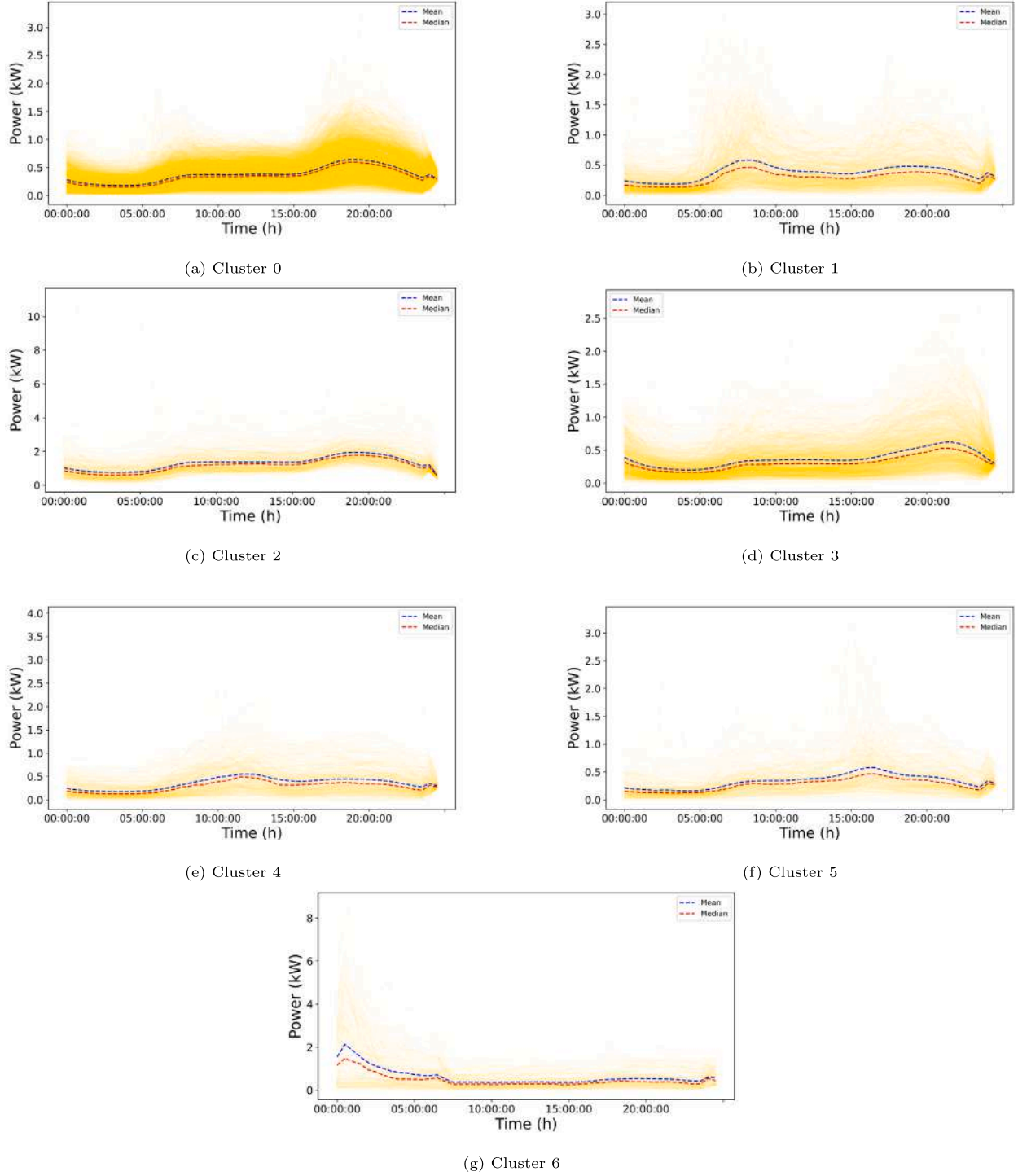


Fig. 6. Cluster visualization.

high Precision, Recall, and F1 Score values, suggesting that it is adept at identifying positive instances with a high degree of accuracy and comprehensively capturing relevant instances. This performance pattern means that it over fits to the existing data, as the algorithm seems to act nearly identically across most classes. This is exactly what we were aiming for, as mentioned earlier. Notably, in Class 4, we observe a deviation from this trend, where all the metric values are slightly lower, hinting at a divergence in algorithm behavior for this particular Class.

As can be seen in Fig. 7, the only classes that hold room for improvement, except three points in class 0, are the ones we mentioned

in our similarity matrices explanation, classes 4 and 5. It is obvious that instances assigned to classes with predictive probabilities falling below the threshold of 0.8 not only appear unsuitable for their designated class but also demonstrate limited suitability for the alternative existing classes.

Another way of validating our results is through xAI [51]. Fig. 8, depicts the average impact of each feature for all the classes for the whole testing dataset as well as Fig. 9 for a random household. This plot offers profound insights, thereby underscoring the substantial practical implications of our methodology in a business context. The association between each class and its respective peak hour is notably clear,

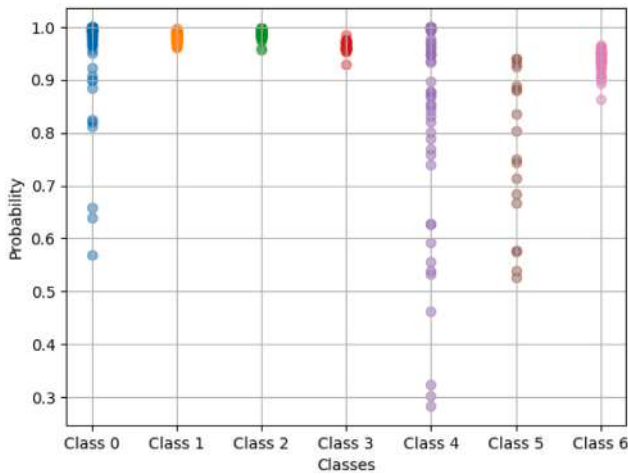


Fig. 7. Probabilities for the predicted classes.

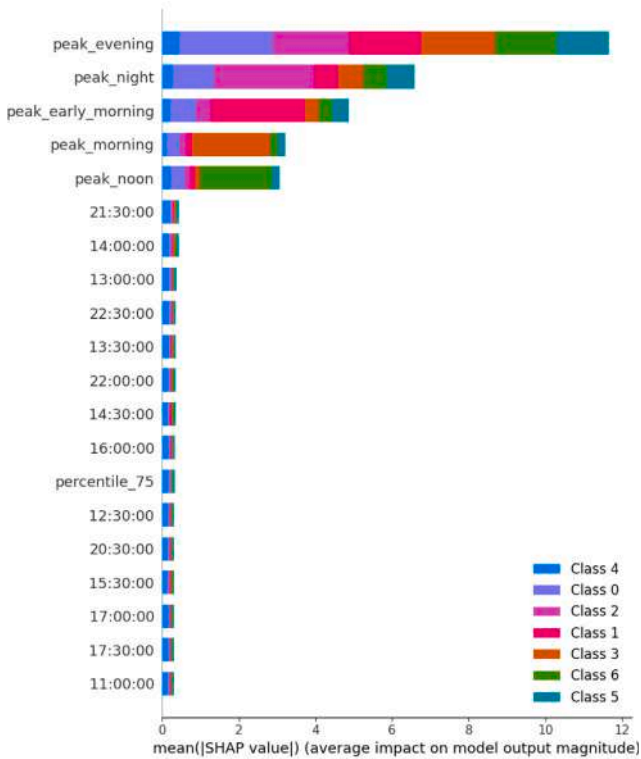


Fig. 8. Average impact of each feature for all classes.

shedding light on the classifier's rationale for grouping them together within a single class.

Furthermore, for programs and initiatives related to data-driven decision-making or targeted interventions, the insights derived from this analysis can serve as a crucial foundation. Specifically, in the case of DR programs, the information gathered from the feature importance analysis can guide the development of more effective and tailored interventions to consumers. By understanding which features have the most significant influence on a user's profile being classified into a specific group, program designers can focus their efforts on addressing those key factors.

In this step having the aforementioned knowledge about the clusters produced, we can restart our pipeline for the user profiles that were assigned to clusters 4 and 5. For simplicity we are going to reach only

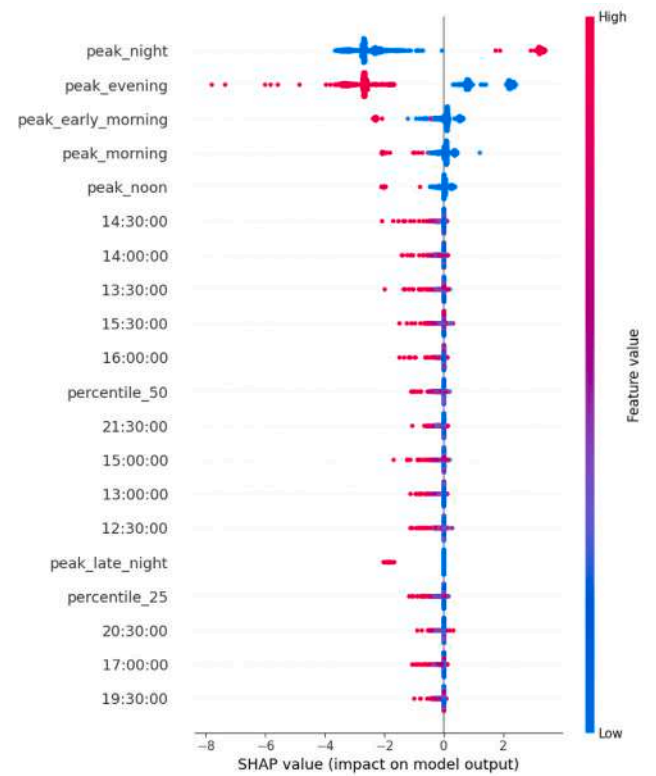


Fig. 9. Average impact of each feature for a random household.

Table 5

Optimal number of clusters for each algorithm based on three evaluation indexes.

Clustering algorithm	SIL	DBI	CHI
K-means	4	4	3
HAC	4	4	4

the second level of our methodology, constructing the clusters in the first place. As we can see in Fig. 10 and in Table 5 those two classes can be split in half when processed alone. Obviously, K-medoids is not a part of this analysis as it was excluded in earlier steps. Following this, the density based algorithm again produces 4 classes for the best indexes when tested for different (ϵ) values.

As previously mentioned, these user profiles do not align with any of the pre-established clusters. As a result of this reevaluation, we choose to exclude these clusters from further analysis due to their evident dissimilarity. It should be noted that these clusters collectively represent a significant portion, approximately 10%, of the entire dataset. This disparity in residential user or household behavior may arise due to discontinuity caused by various factors such as changes in daily routines, lifestyle alterations, external influences like seasonal variations in energy consumption patterns or discrepancies in appliance usage. Their exclusion is imperative to ensure the effectiveness and reliability of DR initiatives, as their erratic behavior could introduce instability and undermine the program's objectives. This discovery and its adaptable nature is a valuable asset in electrical user profiling, as it ensures that our pipeline remains relevant and effective in capturing the inherent complexity of user behavior over time.

As we continue our analysis, it is essential to explore the relationships between the remaining clusters and various DR programs. Without equal treatment, the potential for maximum active demand-side participation that enhances market efficiency is unlikely to materialize. In our case, data required for analyzing Critical Peak Pricing (CPP) or dynamic pricing programs, such as detailed pricing information and

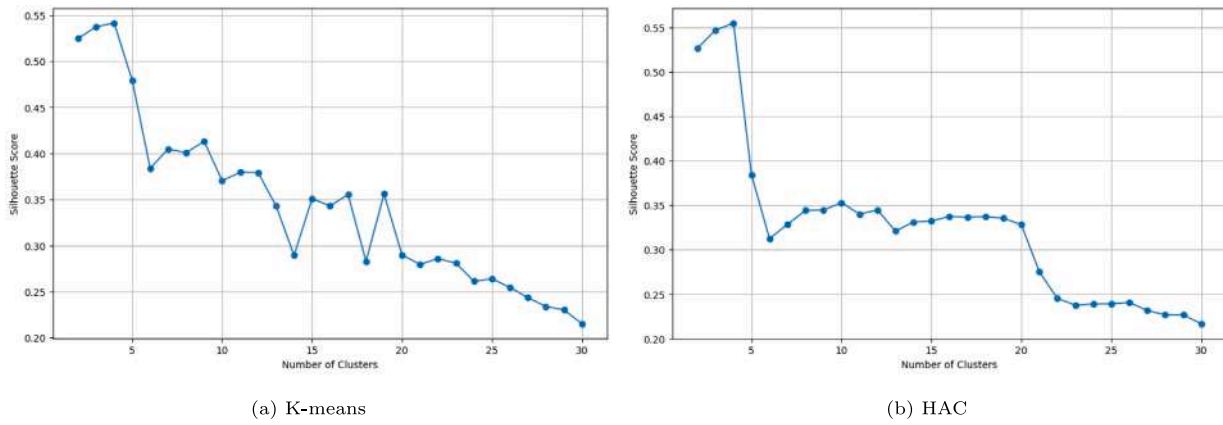


Fig. 10. SIL for every number of clusters and algorithm.

real-time consumption data are absent as we are working solely with historical data. Therefore, ToU programs are the ideal fit for implementation based on the time periods defined in Section 3.1. Supposing we wish to encourage consumers to concentrate their consumption during the evening, a pricing strategy that involves keeping prices low at that time while maintaining higher prices before and after (noon and night) should be implemented. Furthermore, for peak hours, the pricing strategy should be reversed, entailing the establishment of elevated prices during those periods and reduced prices during earlier and later time frames. In this context, as evident from our results, this program finds effective application in clusters 0,1,2 and 3 with higher prices during early morning, noon and night and comparatively lower in the morning, evening and late night. Accordingly, for cluster 6, we would suggest higher prices during late night hours and lower prices during the early morning and night. However, this recommendation is in stark contrast with the proposed DR program, prompting the need for an alternative initiative to be proposed for the specified cluster. By adopting this approach, our findings can serve as a foundation for the creation of more DR programs.

5. Discussion

The timely and accurate classification of electricity consumers into clusters is crucial for optimizing energy distribution and ensuring efficient utilization of resources. Our study, while contributing valuable insights to the energy sector, is not without its limitations. Notably, this research is relied solely on residential data and thus creates uncertainties regarding the reactions of the proposed method when extended to power systems. The scope of residential data may not fully capture the intricacies and dynamics inherent in broader power system scenarios, raising questions about the method's robustness and efficacy in such expansive and varied environments. Moreover, the available data primarily pertain to energy consumption rather than specific demand management programs. Nevertheless, we believe our solution would be sufficient in the case that data of this kind were acquired and would be equally efficient. It is evident that our methodology can output clusters that can be incorporated into tailored DR programs, due to the insights of user behavior retrieved from our analysis.

6. Conclusions and future work

In this paper, we address a critical challenge in the realm of electrical load segmentation and by extension DR programs, which play a pivotal role in achieving sustainable and resilient energy systems. The efficient management of electricity consumption in households is of paramount importance in the contemporary energy landscape, and the segmentation of electricity consumers into distinct groups based on their consumption patterns and behaviors is a fundamental

requirement for the successful implementation of DR programs. Given the lack of mature information systems to support such assessments, several projects are currently straggling to receive financial support by funding aggregators to incorporate smart meters in order to maximize the knowledge of residential electrical load consumption.

In this respect, a data-driven methodology is proposed which aspires to pave the way for aggregators and electricity providers to formulate and adjust DR programs based on their business and environmental goals. The proposed framework includes unsupervised learning techniques, namely clustering algorithms, that based on household electricity trends analysis, can select the optimal number of user groups based on some related indexes. Furthermore, the results are then evaluated through similarity matrices between the algorithms, selecting the nominal groups and the ones that need more attention. Finally, the problem is reincarnated as a probabilistic classification one making way for xAI techniques to give further business insights into the reason why these clusters were created in the first place.

To expound upon this further, it is crucial to delve into the scalability implications of our approach, particularly in light of integrating the classifier component. Scalability concerns can encompass several facets, including the capacity to handle larger datasets, the computational efficiency required for processing, and the adaptability to a broader range of applications and use cases. The methodology is evaluated considering a real case study in London. In conclusion, the clustering algorithm analysis indicates that the ideal number of clusters in this scenario is seven. However, our methodology reveals that approximately 10% of the dataset comprises two clusters with notable internal dissimilarities. As a result, we opt to omit these clusters from further analysis and consideration for DR programs.

Incorporating these insights into DR programs can lead to enhanced user outcomes, cost-effectiveness, and overall program efficiency. It allows for a more precise allocation of resources and interventions to individuals based on the specific factors that influence their classification, ultimately improving the program's ability to achieve its objectives. While acknowledging the limitations of this study, it is imperative that future research in the field of electricity user segmentation places a heightened emphasis on incorporating a comprehensive range of factors. These factors should encompass not only behavioral and socio-demographic aspects but also dig into the intricate realm of psychological factors that influence consumer choices and behaviors. By integrating these dimensions into our segmentation approach, we can construct more completed and meaning full user profiles for DR programs. More importantly, the proposed approach ought to become more relevant for large-scale applications and datasets. The inclusion of real-time data, coupled with robust forecasting capabilities for predicting electricity consumption trends in the coming days, would significantly reduce the complexity and limitations of the current methodology. Moreover, addressing the limitations of relying solely

on residential data and the absence of specific demand management program data would be a huge step following this study. Expanding the scope and incorporating comprehensive data aligned with demand management initiatives as well as going beyond the residential sector will improve the practicality of the proposed method in real-world settings.

CRedit authorship contribution statement

Vasilis Michalakopoulos: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Elissaios Sarmas:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Ioannis Papias:** Writing – review & editing, Writing – original draft. **Panagiotis Skaloumpakas:** Writing – review & editing, Writing – original draft. **Vangelis Marinakis:** Writing – review & editing, Writing – original draft. **Haris Doukas:** Writing – review & editing, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The link for the dataset is within the manuscript.

Acknowledgments

The work presented is based on research conducted within the framework of the Horizon Europe European Commission project DEDALUS (Grant Agreement No. 101103998). The content of the paper is the sole responsibility of its authors and does not necessarily reflect the views of the EC.

References

- [1] Sarmas Elissaios, Spiliotis Evangelos, Marinakis Vangelis, Tzanes Georgios, Kaldellis John K, Doukas Haris. ML-based energy management of water pumping systems for the application of peak shaving in small-scale islands. *Sustainable Cities Soc* 2022;82:103873.
- [2] Albadi MH, El-Saadany EF. Demand response in electricity markets: An overview. In: 2007 IEEE power engineering society general meeting. 2007, p. 1–5.
- [3] Bahrami Sh, Parniani M, Vafaimehr A. A modified approach for residential load scheduling using smart meters. In: IEEE PES innovative smart grid technologies conference Europe. 2012.
- [4] Parvania Masood, Fotuhi-Firuzabad Mahmud, Shahidehpour Mohammad. Optimal demand response aggregation in wholesale electricity markets. *IEEE Trans Smart Grid* 2013;4(4):1957–65, Cited by: 271.
- [5] Muratori Matteo, Rizzoni Giorgio. Residential demand response: Dynamic energy management and time-varying electricity pricing. *IEEE Trans Power Syst* 2016;31(2):1108–17, Cited by: 294.
- [6] Stanelyte Daiva, Radziukyniene Neringa, Radziukynas Virginijus. Overview of demand-response services: A review. *Energies* 2022;15(5).
- [7] Kaur Ramanpreet, Gabrijelčić Dušan. Behavior segmentation of electricity consumption patterns: A cluster analytical approach. *Knowl-Based Syst* 2022;251:109236.
- [8] Qiu Dawei, Wang Yi, Wang Junkai, Jiang Chuanwen, Strbac Goran. Personalized retail pricing design for smart metering consumers in electricity market. *Appl Energy* 2023;348. Cited by: 0; All Open Access, Hybrid Gold Open Access.
- [9] Shi Mengge, Wang Han, Xie Peng, Lyu Cheng, Jian Linni, Jia Youwei. Distributed energy scheduling for integrated energy system clusters with peer-to-peer energy transaction. *IEEE Trans Smart Grid* 2023;14(1):142–56, Cited by: 3.
- [10] Qiu Dawei, Ye Yujian, Papadaskalopoulos Dimitrios, Strbac Goran. Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach. *Appl Energy* 2021;292. Cited by: 53.
- [11] Kwac Jungsuk, Flora June, Rajagopal Ram. Household energy consumption segmentation using hourly data. *IEEE Trans Smart Grid* 2014;5(1):420–30.
- [12] Wang Yi, Chen Qixin, Hong Tao, Kang Chongqing. Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Trans Smart Grid* 2019;10(3):3125–48, Cited by: 673; All Open Access, Green Open Access.
- [13] Motlagh Omid, Berry Adam, O'Neil Lachlan. Clustering of residential electricity customers using load time series. *Appl Energy* 2019;237:11–24.
- [14] Jiang Zigui, Lin Rongheng, Yang Fangchun. A hybrid machine learning model for electricity consumer categorization using smart meter data. *Energies* 2018;11(9).
- [15] Sarmas Elissaios, Stropoulos Sofoklis, Marinakis Vangelis, Santori Francesca, Bucarelli Marco Antonio, Doukas Haris. An incremental learning framework for photovoltaic production and load forecasting in energy microgrids. *Electronics* 2022;11(23):3962.
- [16] Testasecca Tancredi, Lazzaro Marilena, Sarmas Elissaios, Stamatopoulos Stathis. Recent advances on data-driven services for smart energy systems optimization and pro-active management. In: 2023 IEEE international workshop on metrology for living environment. IEEE; 2023, p. 146–51.
- [17] Rajabi Amin, Eskandari Mohsen, Ghadi Mojtaba Jabbari, Li Li, Zhang Jiangfeng, Siano Pierluigi. A comparative study of clustering techniques for electrical load pattern segmentation. *Renew Sustain Energy Rev* 2020;120:109628.
- [18] Chicco Gianfranco. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy* 2012;42(1):68–80, 8th World Energy System Conference, WESC 2010.
- [19] Panapakidis Ioannis P, Alexiadis Minas C, Papagiannis Grigoris K. Deriving the optimal number of clusters in the electricity consumer segmentation procedure. In: 2013 10th international conference on the European energy market. 2013, p. 1–8.
- [20] McLoughlin Fintan, Duffy Aidan, Conlon Michael. A clustering approach to domestic electricity load profile characterisation using smart metering data. *Appl Energy* 2015;141:190–9.
- [21] Wang Yanlong, Li Li, Yang Qinmin. Application of clustering technique to electricity customer classification for load forecasting. In: 2015 IEEE international conference on information and automation. 2015, p. 1425–30.
- [22] Schäfer Hanna, Viegas Joaquim L, Ferreira Marta C, Vieira Susana M, Sousa JMC. Analysing the segmentation of energy consumers using mixed fuzzy clustering. In: 2015 IEEE international conference on fuzzy systems. 2015, p. 1–7.
- [23] Rajabi Amin, Li Li, Zhang Jiangfeng, Zhu Jianguo, Ghavidel Sahand, Ghadi Mojtaba Jabbari. A review on clustering of residential electricity customers and its applications. In: 2017 20th international conference on electrical machines and systems. 2017, p. 1–6.
- [24] Soares João, Lezama Fernando, Pinto Tiago, Morais Hugo. Complex optimization and simulation in power systems. *Complexity* 2018;2018:1–3.
- [25] Yilmaz S, Chambers J, Patel MK. Comparison of clustering approaches for domestic electricity load profile characterisation - implications for demand side management. *Energy* 2019;180:665–77.
- [26] Zarabie Ahmad Khaled, Lashkarbolooki Sahar, Das Sanjoy, Jhala Kumarsinh, Pahwa Anil. Load profile based electricity consumer clustering using affinity propagation. In: 2019 IEEE international conference on electro information technology. EIT, 2019, p. 474–8.
- [27] Al Khafaf Nameer, Jalili Mahdi, Sokolowski Peter. A novel clustering index to find optimal clusters size with application to segmentation of energy consumers. *IEEE Trans Ind Inf* 2021;17(1):346–55.
- [28] Czettány László, Vámos Viktória, Horváth Miklós, Szalay Zsuzsa, Mota-Babiloni Adrián, Deme-Bélafi Zsófia, et al. Development of electricity consumption profiles of residential buildings based on smart meter data clustering. *Energy Build* 2021;252:111376.
- [29] Eskandarnia Elham, Al-Ammal Hesham M, Ksantini Riadh. An embedded deep-clustering-based load profiling framework. *Sustainable Cities Soc* 2022;78:103618.
- [30] Okereke George Emeka, Bali Mohamed Chaker, Okwueze Chisom Nneoma, Ukekwe Emmanuel Chukwudi, Echezona Stephenson Chukwukanedu, Ugwu Celestine Ikechukwu. K-means clustering of electricity consumers using time-domain features from smart meter data. *J Electr Syst Inf Technol* 2023;10(1):1–18.
- [31] Chicco G, Napoli R, Piglion F. Comparisons among clustering techniques for electricity customer classification. *IEEE Trans Power Syst* 2006;21(2):933–40.
- [32] Sarmas Elissaios, Forouli Aikaterini, Marinakis Vangelis, Doukas Haris. Baseline energy modeling for improved measurement and verification through the use of ensemble artificial intelligence models. *Inform Sci* 2024;654:119879.
- [33] Kwac Jungsuk, Tan Chin-Woo, Sintov Nicole, Flora June, Rajagopal Ram. Utility customer segmentation based on smart meter data: Empirical study. In: 2013 IEEE international conference on smart grid communications (smartGridComm). 2013, p. 720–5.
- [34] Devijver Emilie, Goude Yannig, Poggi Jean-Michel. Clustering electricity consumers using high-dimensional regression mixture models. *Appl Stoch Models Bus Ind* 2020;36(1):159–77.
- [35] López José J, Aguado José A, Martín F, Muñoz F, Rodríguez A, Ruiz José E. Hopfield-k-means clustering algorithm: A proposal for the segmentation of electricity customers. *Electr Power Syst Res* 2011;81(2):716–24.
- [36] Batrinu Florentin, Chicco Gianfranco, Napoli Roberto, Piglion Federico, Postolache Petru, Scutariu Mircea, et al. Efficient iterative refinement clustering for electricity customer classification. In: 2005 IEEE Russia power tech. 2005, p. 1–7.

- [37] Wang Xuejin, Zhou Chongdong, Yang Yijing, Yang Yueyong, Ji Tianyao, Wang Jifei, et al. Electricity market customer segmentation based on DBSCAN and k-means : —A case on Yunnan electricity market. In: 2020 Asia energy and electrical engineering symposium. AEEES, 2020, p. 869–74.
- [38] Zhang Liping, Deng Song, Li Shiyue. Analysis of power consumer behavior based on the complementation of K-means and DBSCAN. In: 2017 IEEE conference on energy internet and energy system integration. 2017, p. 1–5.
- [39] Arthur David, Vassilvitskii Sergei. K-means++ the advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms. 2007, p. 1027–35.
- [40] Kaufman Leonard, Rousseeuw Peter J. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons; 2009.
- [41] Ester Martin, Kriegel Hans-Peter, Sander Jorg, Xu Xiaowei, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd, vol. 96, 1996, p. 226–31.
- [42] Rousseeuw Peter J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.
- [43] Davies David, Bouldin Don. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979;PAMI-1:224–7.
- [44] Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat 1974;3(1):1–27.
- [45] Van der Maaten Laurens, Hinton Geoffrey. Visualizing data using t-SNE. J Mach Learn Res 2008;9(11).
- [46] Chen Tianqi, Guestrin Carlos. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: ACM; 2016, p. 785–94.
- [47] Ke Guolin, Meng Qi, Finley Thomas, Wang Taifeng, Chen Wei, Ma Weidong, et al. Lightgbm: A highly efficient gradient boosting decision tree. In: Advances in neural information processing systems, vol. 30, 2017, p. 3146–54.
- [48] Prokhorenkova Liudmila, Gusev Gleb, Vorobev Aleksandr, Dorogush Anna Veronika, Gulin Andrey. CatBoost: Unbiased boosting with categorical features. In: Advances in neural information processing systems, vol. 31, 2018.
- [49] Hancock John T, Khoshgoftaar Taghi M. CatBoost for big data: An interdisciplinary review. J Big Data 2020;7(1):1–45.
- [50] Ghori Khawaja Moyeezullah, Abbasi Rabeeh Ayaz, Awais Muhammad, Imran Muhammad, Ullah Ata, Szathmary Laszlo. Performance analysis of different types of machine learning classifiers for non-technical loss detection. IEEE Access 2020;8:16033–48.
- [51] Panagoulas Dimitrios P, Sarmas Elissaios, Marinakis Vangelis, Virvou Maria, Tsihrintzis George A, Doukas Haris. Intelligent decision support for energy management: A methodology for tailored explainability of artificial intelligence analytics. Electronics 2023;12(21):4430.