From insights to dialogue: consumer segmentation to support active participation in energy services

Ivana Milev[1, 2], Lev Prislan[2], Matej Zajc[1]

[1] *Faculty of Electrical Engineering, University of Ljubljana, Tržaška Cesta 25, 1000 Ljubljana, Slovenia*
[2] *GEN-I, Trgovanje in Prodaja Električne Energije D.O.O., Vrbina 17, 8270 Krško, Slovenia*
E-pošta: ivana.milev.im@gmail.com

Abstract

Encouraging consumers to actively participate in the green transformation requires transforming raw data from smart home equipment and smart meters into tangible information. Machine learning algorithms enable real-time analysis of enormous volumes of data and deliver precise predictions of many types. Clustering consumers allows for better communication with them and to provide them with personalized feedback. Consumers benefit from receiving informative and motivated consumption feedback since the main purpose of consumption feedback is to make electricity visible. This paper presents an example of household clustering for personalized communication. Results show that consumers can be clustered based on their daily consumption profile.

Keywords: Green energy transition, Consumption feedback, clustering, personalized communication, energy consumption, …

1 Introduction

The demand is still rising worldwide. It has climbed by around 20% over the past two decades, and it is projected to expand at a pace of about 1.8% per year until 2030 [1], [2]. By establishing ambitious goals, initiatives, regulatory frameworks, and technological advancements, the European Union is moving closer to a low-carbon future. These include putting consumers first, increasing renewable energy to at least 42.5% of total energy consumption by 2030, limiting global temperature rise to below 2 °C, boosting energy efficiency to address increased electrification, and reducing greenhouse gas emissions by at least 55% by 2030 compared to 1990 [3], [4], [5].

Consumers are recognised as key actors who are accelerating the green energy transition through their use of renewable energy, investments in innovative clean energy technology, energy saving measures, and flexibility [1]. Their importance creates new responsibilities that replace their formerly passive roles. Historically, consumers were traditional, and passive "recipients of technologies"[6], [7], [8]. It was believed that they are just an "isolated node" that needed power from the grid [9]. These "normal and passive electricity buyers" were subsequently replaced by the energy transition with active consumers who are worried about climate change and who take steps to reduce their energy use.

In the age of "smart-everything," consumers are overwhelmed by the constant flow of information. Furthermore, energy is not a conventional consumable item; instead, it is invisible, inconspicuous, abstract, and intangible, unable to be felt, smelled, or touched, and so difficult to explain [10].The transition from daily habits like heating, cooking, and watching TV to a unique energy environment in which neighbours can "borrow electricity" from neighbours or even strangers necessitates a new design approach. One that puts the consumer first, handles the complexity, and ensures that no one falls behind.

Transparency and clarity in information are thought to be powerful motivators and nudges [11]. Providing consumers with instructive and motivating consumption feedback has been shown to be beneficial [12]. The primary goal of consumption feedback is to make electricity visible [13]. Feedback is viewed as a learning tool that enables, informs, and directs consumers [14]. When supplied to consumers on a regular basis over a set period of time, they may become more mindful of their energy consumption [15]. Furthermore, giving consumers information about their energy usage, as well as comparisons or recommendations and instructions, can lead to energy savings [16] and motivate them to adopt sustainable behaviour.

However, general information and tips are not enough. Instead, personalized communication with the consumers about their energy habits is necessary. Finding the most effective way to communicate with consumers involves analysing consumer and environmental data obtained over time using IoT appliances [17]. Processing large datasets in a reasonable period of time is not always supported by traditional statistical approaches. Cognitive or machine learning algorithms allow for real-time analysis of massive amounts of data and provide precise predictions of many types, such as product recommendations, consumer segmentation, fraud and risk detection, client retention, and so on [18].

Segmenting consumers provides advantages to the main actors in the energy domain: a) energy suppliers - alows for personalized communication with their customers and improved design of new energy services; b) consumers – improved understanding of own energy behaviour and energy savings; and c) the environment – optimal utilization of the grid will support the fight against climate change [19].

This paper presents an example of household clustering for personalized communication. The electricity data was collected with the smart meters, then processed with Microsoft SQL Management Studio program. Households were than clustered using K-Medoid clustering algorithm in KNIME program. The results were next analyzed in Power BI program.

2 Methodology

The demonstrator site was established as part of a H2020 funded project [20]. Around 10,000 Slovenian households were invited to participate in the study and around 1,082 accepted the invitation and agreed to share their data. All of the households were equipped with smart meter but the data was incomplete for some of them. Therefore, we used the data from the smart meters of 824 households for one year in total to cluster them based on their daily consumption habits, following the methodology as presented in Figure 1.



Figure 1. Methodology for household clustering.

a) Data collection

The electricity consumption data - energy in kWh on 15- minute interval, was *collected* by the smart meters and transfered to the EVT - Enotna Vstopna Točka (Single point of entry) database and later to the DSO and the Utility which was part of the project. The data was actually timeseries that included the metering point ID, the measurement date and time, and the amount of energy used over a 15-minute period.

b) Data processng

The data was than *processed* with the Microsoft SQL Management Studio program for further analysis. During this step, the households with missing data were filtered out and new features were produced. First the energy consumption was aggregated for each household on hourly basis for the whole year, than normalized consumption was prepared – percentage of hour energy consumption regarding overall energy consumption, as presented in Table 1. This was neccessary for the next step, where clustering will be used to cluster households.

Table 1. Example of electricity data after processing

Household	Hour	Average hourly energy consumption [%]
Household 1	0	4.15
Household 1	1	3.75
Household 1	2	3.67

c) Clustering

The last step was to prepare the clusters using clustering techiques. Cluster analysis organises data items based on information in the data that characterises the objects and their relationships, with the purpose of ensuring that objects within a cluster are similar (or linked) to one another while differing from objects in other clusters [21]. The major clustering approaches are classified as partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods [22]. K-medoid, or partitioning around Medoids, is a technique that seeks to find *k* partitions for *n* objects. A medoid serves as a reference point for a generally centrally positioned object in a cluster- the cluster heads for each cluster [23]. Following an initial random selection of *k* medoids, the algorithm iteratively attempts to make a better choice of medoids [22]. After that, the cluster is formed based on the distance between the data points [23].

We are looking for "typical households" among the sample. That is why the K-medoids clustering algorithm was selected. We selected three clusters of households, which is the number of partitions *k*, among 824 households which is the number of objects *n*. The number of cluster *k* =3 was selected after running the model several times searching for 3, 4, 5 and 6 clusters and analysing the results: average cluster distances, energy consumption profiles for the selected clusters and the size of the clustes.

d) Implementation in Knime

After the cluster heads are defined, the algorithm assigns the rest of the households to one partition based on the distance between data points, forming the clusters. The program used for the clustering process was KNIME. KNIME's modular data pipeline architecture integrates a variety of machine learning and data mining components. A graphical user interface and the usage of JDBC enable the creation of nodes that combine many data sources, including nodes for preprocessing, modelling, data analysis, and visualisation. The clustering model in Knime, used for the clustering of households is presented in Figure 2.

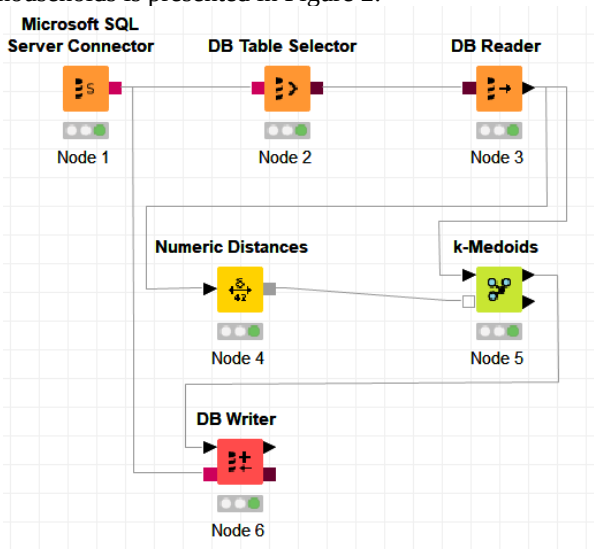


Figure 2. Clustering model in Knime.

First step is to connect to the database to load the dataset into the program, which is done by Nodes 1-3. Node 1 establishes a connection with a Microsoft SQL Server. There are no input ports here. The parameters that must be setup include the server's hostname (or IP address), port, database name, and login credentials. The output, representing the connection to the server and database, is then connected to node number 2 – to select a table. Node 3 than executes the input query in the database and retrieves the result into a KNIME data table.

The clustering process is then executed with Nodes 4 and 5. Node 4 computes the distances across numerical columns. This node's input is the dataset retrieved from the database. The parameters that needed configuration are: list of the columns which will be used during the algorithm, type of distance that need to be calculated and how to handle the missing values. Regarding distance metrics, the Euclidean distance was used.

The actual performance of the clustering with K-Medoids algorithm was done with Node 5. This node uses the dataset retrieved with Node 3 and the distances calculated with Node 4. It begins with a random initialization of the medoids and then iteratively does an exhaustive search on the input data, calculating the cost of swapping any medoid with any input data row. The number of clusters and chunk size must be specified here. Once the cluster heads are defined and all of the households are assigned to the closest cluster head, forming the clusters, the results are then exported to the database with Node 6. The results are then processed with Microsoft SQL Management Studio program and analysed with Microsoft Power BI program.

3 Results

An example of the clustering results in Knime is presented in Figure 3. Column »UserID« present the ID of the households, »Row584«, »Row223« and »Row507« represent the distance of the household to all of the clusters and column »Cluster« represent the assigned cluster which is later renamed to Cluster 1, Cluster 2 and Cluster 3. For example, UserID 818 is closest to Row223 which is why this household is assigned to cluster Row223.

The number of households assigned to each cluster as well as the average distance to all of the clusters is presented in Table 2. Results show that Cluster C2 is the largest one with 387 of the households (47%), followed by Cluster C2 with 244 households (30%) and C1 with 193 households (23%).

USERID	Row584	Row223	Row507	Cluster
225				Row584
334				Row223
648				Row507
716				Row584
818				Row223
886				Row507
944				Row507
975				Row223
999				Row584

Figure 3. Example of the clustering results in Knime.

The greater the difference inside the cluster means also greater heterogeneity between households. The average distance within the cluster is expected to be the shortest, which it is in this case. For example, when we compare the average distance for cluster number one to all others (including cluster number one itself), we can see that the average distance within the cluster is 0.256, which is the shortest. It's the same for the other clusters. The distribution of distances to own cluster among households is presented in Figure 4.

Table 2. Cluster size and distances

ID	DISTANCES			SIZE (HOUSEHOLDS)
	C1	C2	C3	
C1	**0.26**	0.33	0.41	193
C2	0.36	**0.25**	0.39	387
C3	0.41	0.37	**0.22**	244

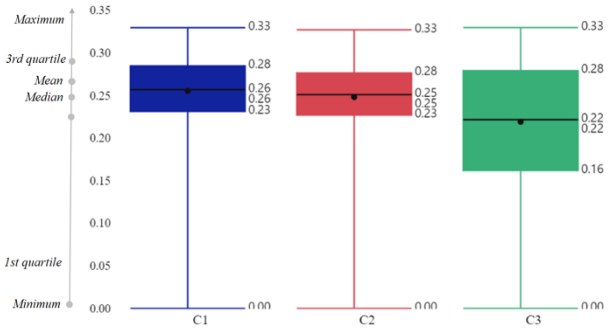


Figure 4. Distance to own cluster among households.

The electricity data shows that on average one household consumes around 5,988 kWh per year, 499 kWh per month or 16.41 kWh per day. In fact, one household from Cluster 1 on average consumes around 4,441 kWh per year, Cluster 2 around 4,940 kWh per year and Cluster 3 around 8,873 kWh per year. Average daily electricity consumption of each cluster is presented in Figure 5, showing the difference among clusters.

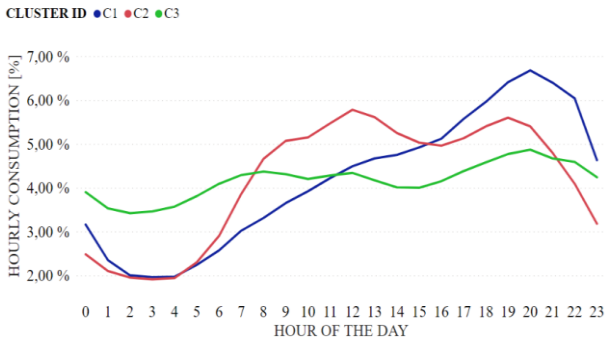


Figure 5. Daily consumption [%] by hours and clusters.

Electricity consumption data show that households from Cluster 1 have increasing consumption throughout the day with an evening peak. Households from Cluster 2 have two peaks: morning and evening while households from Cluster 3 have more constant consumption throughout the day.

4 Discussion and conclusion

The Green energy transition puts consumers in the center. Consumers are not longer seen as passive buyers of electricity, but as the catalysts of the energy transition. They are increasingly regarded as active consumers, prosumers and flexumers. In the past, consumers received basic information about their energy consumption and the payment data which was later broadened and enchaned with new information. Electronic bills ultimately replaced paper bills and they have later expanded into energy platforms and applications [1]. Energy platforms serve as a link between consumers and their smart meters or other IoT devices in their homes [24].

However, in an age of information overload, consumers may feel overwhelmed. Furthermore, electricity is an invisible and untypical consumable good which additinally brings new challenges about consumer awareness, engagement and involvement. Feedback increases consumer awareness, allows them to make more informed decisions and leads to increased engagement with energy services. Active consumers are more likely to begin adopting new energy services.

Clustering consumers based on similar characteristics allows for better understanding of their preferences and needs as well as for focused communication with them ensuring that each consumers or cluster receives relevant feedback. Consumer segmentation with clustering or other ML algorithms is an effective method to enable personalized communication with consumers. Clustering allows to identify consumer behaviour that would not be noticeable through regular data analysis.

The results of our clustering analysis has shown that all three clusters of households have own energy behaviour and daily profile which can be further used for personalized comunication with them in terms of time and place. For example, based on the constant electricity consumption throughout the day and on their high annual electricity consumption – nearly three times the average Slovenian household, we could say that households from Cluster 3 are using electricity and probably a heat pump for space heating. This information could be used to sent personalized consumption feedback to this cluster of consumers which could have great impact on their energy behavior. Space heating is the most intensive energy-related activity in terms of consumption, with around 60% of the total household energy consumption and any reductions in this field could support the energy transition. Another example, households from Cluster 1 have one energy consumption peak during the day – in the evening. This information could be used to choose the most suitable time of feedback delivery – when they are or are not at home, so that the feedback information could have the most impact on them. Furthermore, this information could help to choose the most suitable feedback delivery channel: at the place of consumption when consumers are at home or through energy platform and mobile apps when they are not, etc.

Acknowledgment

The data set was collected within the Newcomers project, funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 837752, following the data management plan for the collection, storage and processing of personal data. Author M. Zajc was supported by ARIS under the Scientific Program P2-0246.

Literature

[1] I. Milev, L. Prislan, and M. Zajc, “Energy Portal Design and Evaluation for Consumer Active Participation in Energy Services : Seven-Month Field Study with 234 Slovenian Households,” pp. 1–21, 2022.

[2] C. Cruz, T. Alskaif, E. Palomar, and I. Bravo, “Prosumers integration in aggregated demand response systems,” *Energy Policy*, vol. 182, no. April, p. 113745, 2023, doi: 10.1016/j.enpol.2023.113745.

[3] G. S. Seck *et al.*, “Hydrogen and the decarbonization of the energy system in europe in 2050: A detailed model-based analysis,” *Renewable and Sustainable Energy Reviews*, vol. 167, no. June 2022, p. 112779, 2022, doi: 10.1016/j.rser.2022.112779.

[4] D. Sica, B. Esposito, S. Supino, O. Malandrino, and M. R. Sessa, “Biogas-based systems: An opportunity towards a post-fossil and circular economy perspective in Italy,” *Energy Policy*, vol. 182, no. July, p. 113719, 2023, doi: 10.1016/j.enpol.2023.113719.

[5] H. Brugger, W. Eichhammer, N. Mikova, and E. Dönitz, “Energy Efficiency Vision 2050: How will new societal trends influence future energy demand in the European countries?,” *Energy Policy*, vol. 152, 2021, doi: 10.1016/j.enpol.2021.112216.

[6] I. Gomes, K. Bot, M. da G. Ruano, and A. Ruano, “Recent Techniques Used in Home Energy Management Systems: A Review,” *Energies (Basel)*, vol. 15, no. 8, 2022, doi: 10.3390/en15082866.

[7] M. Kubli, M. Loock, and R. Wüstenhagen, “The flexible prosumer: Measuring the willingness to co-create distributed flexibility,” *Energy Policy*, vol. 114, no. December 2017, pp. 540–548, 2018, doi: 10.1016/j.enpol.2017.12.044.

[8] R. Shortall, A. Mengolini, and F. Gangale, “Citizen Engagement in EU Collective Action Energy Projects,” *Sustainability (Switzerland)*, vol. 14, no. 10, pp. 1–24, 2022, doi: 10.3390/su14105949.

[9] M. Goulden, A. Spence, J. Wardman, and C. Leygue, “Differentiating ‘the user’ in DSR: Developing demand side response in advanced economies,” *Energy Policy*, vol. 122, pp. 176–185, Nov. 2018, doi: 10.1016/J.ENPOL.2018.07.013.

[10] I. Milev and M. Zajc, “Tangible information for active consumers: data from smart home device and smart meter become customer newsletters,” *2022 30th Telecommunications Forum, TELFOR 2022 - Proceedings*, pp. 1–4, 2022, doi: 10.1109/TELFOR56187.2022.9983708.

[11] D. Brounen, N. Kok, and J. M. Quigley, “Energy literacy, awareness, and conservation behavior of residential households,” *Energy Econ*, vol. 38, pp. 42–50, 2013, doi: 10.1016/j.eneco.2013.02.008.

[12] J. Froehlich, “Promoting Energy Efficient Behaviors in the Home through Feedback : The Role of Human-Computer Interaction,” *Computing Systems*, vol. 9, pp. 1–10, 2009.

[13] A. Podgornik, B. Sucic, and B. Blazic, “Effects of customized consumption feedback on energy efficient behaviour in low-income households,” *J Clean Prod*, vol. 130, pp. 25–34, 2016, doi: 10.1016/j.jclepro.2016.02.009.

[14] S. Darby, “THE EFFECTIVENESS OF FEEDBACK ON ENERGY CONSUMPTION,” *Orthod Fr*, vol. 86, no. 3, pp. 221–231, 2006, doi: 10.1051/orthodfr/2015025.

[15] A. Selvefors, I. C. M. Karlsson, and U. Rahe, “5th International Congress of International Association of Societies of Design Research, Consilence and Innovation in Design - Proceedings and Program 5th IASDR 2013 TOKYO, Use and Adoption of Interactive Energy Feedback Systems. Available online: https:/,” *Consilience and Innovation in Design*, pp. 1771–1782, 2013.

[16] M. Shen, Y. Lu, K. H. Wei, and Q. Cui, “Prediction of household electricity consumption and effectiveness of concerted intervention strategies based on occupant behaviour and personality traits,” *Renewable and Sustainable Energy Reviews*, vol. 127, no. March, p. 109839, 2020, doi: 10.1016/j.rser.2020.109839.

[17] M. Elsisi, M. Amer, A. Dababat, and C. L. Su, “A comprehensive review of machine learning and IoT solutions for demand side energy management, conservation, and resilient operation,” *Energy*, vol. 281, no. March, p. 128256, 2023, doi: 10.1016/j.energy.2023.128256.

[18] CogNub, “Cognitive computing and ML”, [Online]. Available: http://www.cognub.com/index.php/cognitive-platform/

[19] F. Rodríguez-Gómez, J. del Campo-Ávila, and L. Mora-López, “A novel clustering based method for characterizing household electricity consumption profiles,” *Eng Appl Artif Intell*, vol. 129, no. August 2023, p. 107653, 2024, doi: 10.1016/j.engappai.2023.107653.

[20] Newcomers, “NEWCOMERS, Cordis. Available: https://cordis.europa.eu/project/id/837752 (accessed on 19 May 2024).” [Online]. Available: https://cordis.europa.eu/project/id/837752

[21] Arman, “Chapter 1 : Introduction Chapter 1 : Introduction,” *Fluid Mechanics*, pp. 1–16, 2018.

[22] A. Batra, “Analysis and Approach: K-Means and K-Medoids Data Mining Algorithms,” *5th IEEE International Conference on Advanced*, no. 274, pp. 274–279, 2011, [Online]. Available: http://www.apiit.edu.in/downloads/all chapters/CHAPTER-59.pdf%0Ahttp://www.icacct.apiit.edu.in/download/all chapters/CHAPTER-59.pdf

[23] G. Surya Narayana and D. Vasumathi, “An Attributes Similarity-Based K-Medoids Clustering Technique in Data Mining,” *Arab J Sci Eng*, vol. 43, no. 8, pp. 3979–3992, 2018, doi: 10.1007/s13369-017-2761-2.

[24] I. Milev, L. Prislan, and M. Zajc, “(UNDER REVIEW) A step closer towards active and flexible consumers: consumption visualization feedback by appliance to increase households’ involvement,” *Environ Res Commun*, 2024.