Exploring Residential Energy Behaviors: Identifying Key Factors and Perspectives for Optimization

Kévin Tang, Mohamed-Essaid Khanouche, and Parisa Ghodous

CNRS, INSA Lyon, Universite Claude Bernard Lyon 1, LIRIS, UMR5205, 69622 Villeurbanne Cedex, France lncs@liris.cnrs.fr
https://liris.cnrs.fr/

Abstract. Household energy consumption represents a significant share of overall energy consumption. To deepen our understanding of surplus energy production, management, and storage, it is crucial to elucidate the implicit patterns of consumer behaviors and identify factors influencing their performance. This paper aims to analyze the household energy consumption pattern, with a focus on the impact of user behavior on energy consumption. To this end, we focus on selecting the most relevant feature subset from a large dataset, which allows for better insights, reduced computation time, and improved predictive performance. The results of this study can help policymakers examine household behaviors at various levels of society to ensure that their quality of life is adequate or needs improvement.

Keywords: Feature selection · Smart home · Household energy consumption · Correlation analysis · User behavior · Residential energy efficiency

1 Introduction

The world population reached 8.2 billion in 2024 [2], one billion more than in 2010, and is expected to cross the 10 billion mark before the end of the century according to United Nations forecasts [3,4].This rapid growth, combined with urbanization and increasing energy needs, is contributing to a significant increase in household energy consumption, which represents about 25% of global energy consumption [5]. In metropolitan France, households currently account for 36% of electricity consumption, making this sector the most energy-intensive in proportion, ahead of the tertiary and industrial sectors [6]. This trend raises significant concerns about energy sustainability, particularly in terms of the increasing strain on the electricity grid, rising carbon emissions, the challenges of transitioning to renewable energy sources, and the need to balance electricity production with growing demand.

The rise of smart Internet of Things (IoT) devices, whose number is expected to reach 50 billion units by 2025 and 100 billion by 2030 [7], opens up new perspectives for monitoring and optimizing energy consumption. These devices collect massive amounts of data on user habits, preferences and behaviors, offering an unprecedented opportunity to understand and manage energy consumption more efficiently. The Internet of Behavior (IoB) is a new paradigm aimed at analyzing data related to user behavioral trends to achieve specific objectives [8,9,10]. This data, collected by smart devices, reflects user behavior, habits, and lifestyle across various domains such as health [11], transportation [12,13], education [14,15], and energy consumption [25,26]. IoB enables the development of energy management models within residential spaces by observing how occupants' behavior influences energy consumption. By continuously monitoring user behavior, IoB provides valuable insights into how energy consumption can be optimized. However, the challenge remains: how to interpret these data to develop efficient energy management models in residential spaces, based on occupant behavior? While existing approaches mainly focus on selecting features related to the devices used [17,18] or the homes themselves (surface area, number of bedrooms) [24], our approach is distinguished by a focus on behavioral features. The way household members use their appliances, such as the frequency of use and the mode of operation, can greatly influence overall energy consumption. For example, excessive use of appliances like washing machines, dryers, or air conditioning systems can lead to significantly higher energy bills. Understanding these usage patterns enables us to identify opportunities for optimization, such as reducing the frequency of use, switching to energy-saving modes. We seek to understand how human behaviors influence energy consumption, thus offering a more detailed analysis and an opportunity to improve overall energy efficiency.

In this paper, we analyze the RECS2020 dataset, which contains information on residential energy consumption of US households in 2020. We develop learning models to identify correlations between user behaviors and their energy expenditures in residential spaces. Our methodology includes collecting and preprocessing behavioral data, selecting relevant features, and applying clustering algorithms to segment households into meaningful groups. These clusters are then analyzed to extract information on consumption habits.

The remainder of this article is structured as follows: In Section 2, we review the related work in the field, providing context for our study. Section 3 presents the problem definition, outlining the challenges and objectives of our research. Section 4 details the methodology, including data preprocessing, clustering techniques, correlation analysis, and data exploration. In Section 5, we discuss the result analysis, where we examine the findings and their implications. Finally, Section 6 concludes the article, summarizing the key insights and potential avenues for future research.

2 Related work

In this section, we review related work in three main areas: energy consumption prediction, feature extraction methods, and techniques for influencing energy use behaviors.

Prediction models are essential tools in understanding and forecasting energy consumption patterns. These models leverage historical data and various influencing factors to estimate future energy usage, enabling more efficient energy management. Mouna Labiadh [24] presents a novel methodology for energy consumption prediction in buildings, especially where historical data is lacking. This approach utilizes a Siamese MLP model combined with KNN and various temporal models like SVR, MLP, LSTM, CNN, and Seq2seq, focusing primarily on the physical characteristics of buildings. Wang et al. [27] developed a deep convolutional neural network (DCNN) based on ResNet, integrating temporal and meteorological features for accurate electricity consumption forecasting. Chan and Yeo [28] employed a Sparse Transformer, which provides comparable accuracy to RNNs while being faster. Hadjout et al. [29,30] used a combination of LSTM, GRU, TCN, and SARIMA models, enhanced by Ensemble Learning techniques for electricity consumption prediction. Syed et al. [31] explored various LSTM architectures, such as bidirectional and unidirectional layers, to forecast electricity consumption. Alhussein et al. [32] integrated hybrid architectures that combine CNN and LSTM for better forecasting accuracy. Li et al. [33] developed the Trans-T2V model, combining a Transformer with Time2Vec for refined electricity consumption forecasts.

Several studies have focused on identifying key factors that influence energy consumption. For instance, Sani et al. [16] analyzed U.S. household energy consumption using the RECS 2015 dataset, but primarily focused on socio-economic factors rather than behavioral aspects. Sanquist et al. [18] identified lifestyle-related factors such as air conditioning usage, laundry practices, and personal computer usage, based on RECS data from 2001 and 2005. Like our study, Heinrich et al. [17] aimed to build behavioral archetypes to better understand energy consumption habits in the residential sector. The data, drawn from the ENERGIHAB project of the French National Research Agency, included 35 variables covering hygiene, food, heating, lighting, leisure practices, and housing occupation from 1363 households in Île-deFrance. This study successfully identified seven distinct behavioral archetypes.

Recent advancements have explored ways to influence and modify energy consumption behaviors. Haya Elayan et al. [25] proposed a decentralized IoB framework aimed at predicting electricity consumption and influencing IoT device behavior. This framework, using French household electricity consumption data and LSTM models, shows promise in reducing energy consumption by directly affecting the operation of connected devices. Although this framework has been tested on a limited number of devices, it offers an optimistic perspective on the possibility of reducing electricity consumption not only at the household level, but also at that of an entire population. Another study by Elayan et al. [26] combined IoB with explainable AI (XAI) to optimize energy use decisions while making these decisions transparent to users. The integration of an Energy Monitor & Controller (EMC) and an "Explainer" component helps users understand and accept energy-saving recommendations, encouraging more sustainable behavior. A well-explained decision, which clearly shows the benefits in terms of energy savings and cost reduction, has the potential to motivate users to adopt more sustainable practices.

The existing literature offers a variety of advanced techniques for energy consumption prediction, feature extraction, and influencing the behaviors of connected devices. Predictive models, such as those based on LSTM [24,29,30] or other neural networks [27,28,31,32], are already well developed to anticipate energy needs. Similarly, sophisticated feature extraction methods allow to analyze socio-demographic factors and building characteristics influencing consumption. Influence techniques also show promising potential to modify user behaviors and optimize energy use. However, this study stands out by focusing specifically on the identification and segmentation of household behaviors, an aspect often underexplored in previous works.Many prior studies were limited by geographic scope, a restricted number of variables, or biases in data collection. Some relied on outdated data, had missing information, or did not focus on user behavior, which constrained their ability to provide a comprehensive analysis of energy consumption patterns. By combining these existing techniques with in-depth behavioral analysis, we are planning to improve the effectiveness of energy interventions. Our approach aims to reduce energy consumption more effectively by targeting specific daily habits and adapting solutions to both occupant behavior and lifestyle patterns. This originality allows us to better understand how behaviors can be modified to reduce the energy footprint, thus contributing to more efficient and targeted energy policies.

3 Problem definition

Managing household energy consumption is a major challenge in the current context of climate change and the transition to renewable energy sources. A better understanding of household energy consumption behaviors can help identify levers for action to reduce energy demand, improve energy efficiency, and decrease greenhouse gas emissions.

The problem lies at the intersection of several issues: environmental, economic, and social. Household energy consumption contributes significantly to greenhouse gas emissions, which has a direct impact on climate change. Reducing energy consumption is essential to achieve the emission reduction targets set by various international agreements. On the economic level, energy represents a significant share of household expenditure. By identifying energy-intensive behaviors and proposing solutions to reduce them, it is possible to reduce household energy bills, which can have a positive impact on purchasing power. On the social level, there are significant disparities in energy consumption based on income, housing size and household equipment. Understanding these disparities is crucial to propose equitable and adapted measures for each type of household.

The main objective of this study is to analyze the influence of behavior on energy consumption. This involves examining how various behaviors impact energy use, identifying distinct consumption patterns, and understanding the key factors driving these patterns. By conducting a thorough analysis of household data, including descriptive statistics and clustering, we aim to uncover the relationship between behavior and energy consumption. The goal is to provide insights into how different behaviors contribute to energy use, which allow us to offer targeted recommendations for reducing consumption based on these findings.

4 Methodology

The objective of this study is to highlight the user behaviors that have the greatest impact on household energy consumption, based on the 2020 RECS dataset. We have selected three target variables: the total amount of electricity used in kilowatt-hours (kWh), the total amount of electricity used in BTU (BTUEL), and the total cost of electricity consumed in dollars (DOLLAREL).

In this section, we first describe the dataset used. Then, we detail the different preprocessings applied to prepare the data for analysis, as well as the feature selection methods employed to identify the most significant behaviors that impact energy consumption. Finally, we analyze the results obtained after the application of clustering algorithms and studying the correlations that allowed us to identify the characteristics that have the greatest influence on the energy consumed.

4.1 Dataset Overview

The Residential Energy Consumption Survey (RECS) is a comprehensive survey conducted in the United States to collect information on the residential energy consumption of American households. Conducted by the Energy Information Administration (EIA), this survey is one of the most comprehensive and reliable sources of data on energy use in the residential sector [1].

The 2020 edition of the RECS collects data on various aspects of household energy consumption, including the types and quantities of fuels used such as electricity, natural gas, propane, and fuel oil, dwelling characteristics, household appliances, heating and cooling systems, as well as recent behaviors such as the use of solar energy or the location of electric vehicle charging. The survey thus covers a wide range of 799 variables, allowing for an in-depth and multidimensional analysis of energy consumption habits. The RECS uses a representative sample of 18496 American households to ensure that the results can be generalized to the entire population. Detailed surveys are sent to selected households and can be administered as paper questionnaires, online, by telephone interviews or during on-site visits.

4.2 Preprocessing

To ensure the quality and relevance of the analyses, several preprocessing steps were applied to the RECS 2020 data. These steps are crucial to clean and prepare the data before applying feature selection techniques and clustering algorithms.

Removal of variables concerning energies other than electricity Since our research aims to identify user behaviors influencing electricity consumption, we removed 117 variables related to other types of energy, such as natural gas, propane, fuel oil and wood. By excluding these variables, we were able to focus our analysis specifically on the factors influencing electricity consumption. This approach reduces the complexity of the dataset and facilitates the interpretation of the results. After this removal, we retained a total of 682 variables.

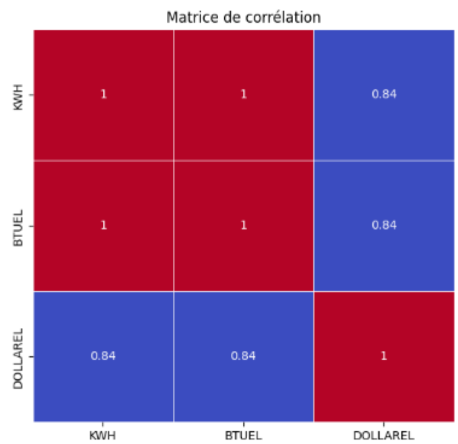
Removal of imputation and calibration indicator variables Imputation indicator variables show whether the values for other variables in the dataset have been imputed, meaning they represent estimates rather than actual data. Calibration variables adjust the weights of responses to ensure the sample accurately reflects the broader population. To refine our analysis and ensure data relevance, we removed 407 of these variables used by the EIA. While these variables are crucial to the EIA's internal processes, they are not necessary for our study, which focuses specifically on electricity consumption behaviors. After these removals, 275 variables remain for our analysis.

Encoding Categorical Variables To prepare our dataset for further analysis, we converted the categorical variables into a form that machine learning algorithms can understand. We used the LabelEncoder method from the scikit learn library. This technique transforms each categorical value into a unique numeric value, indicating the category it belongs to. In our dataset, we identified 7 categorical variables that required encoding. Applying the LabelEncoder to these variables allowed us to convert the categorical data into a numeric representation while preserving the integrity and meaning of the original information.

Missing Values Management To ensure the quality and completeness of our dataset, we adopted a pragmatic approach to handling missing values. The first step was to identify all rows in the dataset that contained one or more missing values. We identified 290 samples containing a NaN value , which represents approximately 1.57% of the dataset. Since this proportion is very small, we decided to remove these samples to simplify data processing. After removing rows with missing values, the number of households in our dataset decreased from 18,496 to 18,206. This approach ensures that our analysis is based on a complete dataset, thereby minimizing potential biases and errors related to imputation of missing values.

Reduction of target variables At the beginning of our study, we identified three main target variables related to electricity consumption: the total amount of electricity used in kilowatt hours (kWh), the total amount of electricity used in BTUs (BTUEL), and the total cost of electricity consumed in dollars (DOLLAREL). However, after performing a Pearson correlation analysis between these variables, we found that some of them were highly correlated. In particular, the

BTUEL variable was highly correlated with kWh, because these two measures are interconvertible (1 kWh is equivalent to approximately 3,412 BTUs). As a result, we decided to reduce the number of target variables to two: kWh and DOLLAREL. This reduction simplifies the analysis while retaining the essential measures to assess electricity consumption. Following this reduction, we have added these 2 variables to our dataset total, bringing the number of variables to 277.



Reducing the number of variables As part of data preprocessing, we performed a Pearson correlation analysis to identify highly correlated variable pairs in our dataset and reduce redundancy. This analysis led to the removal of 69 variables. For example, we removed the variable STATE_POSTAL and kept STATE_NAME to avoid redundancy between state zip codes and their names. Similarly, we removed the variable TELLWORK, which indicates that a household member teleworks, in favor of more precise variables such as TELLDAYS, TLDESKTOP or TLLAPTOP, which provide more information on telework arrangements. This approach was also applied to other domains, such as energy assistance (PAYHELP), electric vehicles (ELECVEH) or housing type (STUDIO), where we chose to keep the variables providing the most detailed and relevant information. The correlation matrix 1 below illustrates the highly correlated variables and supports our decisions for variable reduction by showing how some variables are closely related. After this reduction process, 69 variables were removed, leaving a total of 208 variables for further analysis.

Reducing the number of samples Isolated data, often considered as outliers, can bias machine learning models and distort the results of statistical analyses. It is therefore essential to identify and remove them to obtain more accurate

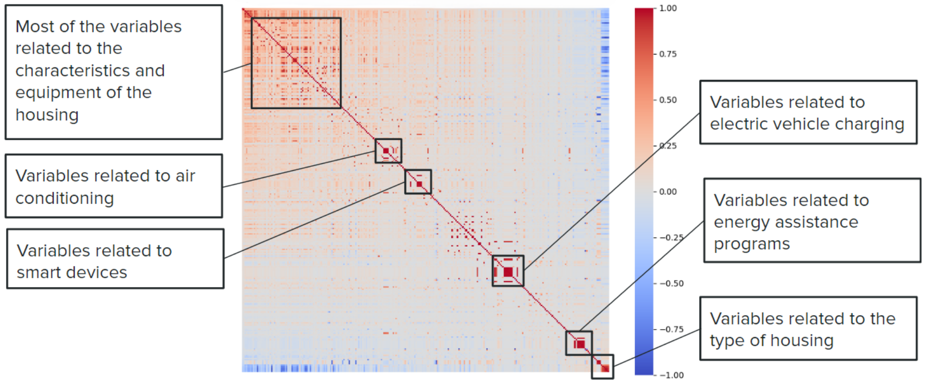


Fig. 1. Correlation matrix of all variables

and reliable analyses, thus allowing a better understanding of the data. In our study, we used the Isolation Forest algorithm [44] to identify these isolated data. Isolation Forest isolates anomalies by building isolation trees. Each tree randomly partitions the data until each point is isolated in a leaf. Anomalies, being rare and different, are isolated more quickly, meaning that they require fewer splits to be separated from other points. Each data point is assigned an anomaly score based on the average depth at which it is isolated in the trees. The shallower the depth, the more likely the point is to be an anomaly. This method detected 894 samples considered as outliers, representing approximately 5% of the dataset. Removing these points reduced the total number of households in our dataset from 18,206 to 17,312. By removing these isolated data points, we significantly improved the quality of the dataset, allowing for more accurate and robust analyses.

Preprocessing Conclusion Following rigorous preprocessing steps, we significantly refined our dataset to make it more relevant and usable for our analysis. The number of samples was reduced from 18,496 to 17,312, following the removal of 894 outliers identified as isolated and 290 samples containing at least one missing value. This reduction allowed us to maintain a more coherent and representative database. Regarding variables, we simplified our dataset by reducing the total number of variables from 799 to 208. This reduction was achieved by eliminating variables related to types of energy other than electricity, imputation and calibration indicator variables, as well as those that were highly correlated. This approach reduced redundancies and focused our analysis on the most relevant factors. These preprocessing steps not only optimized data quality, but also facilitated more precise and meaningful analyses. Ultimately, these adjustments help to obtain more reliable results and deepen our understanding of electricity consumption behaviors.

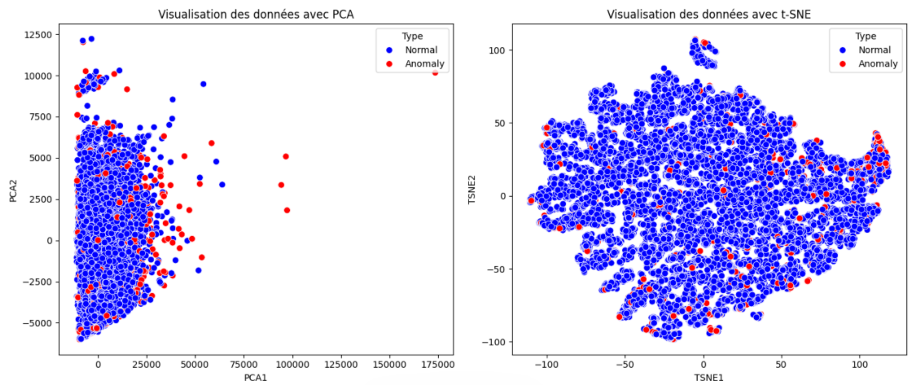


Fig. 2. Visualization of isolated data

Table 1. Summary table of preprocessing steps

Preprocessing steps	Number of variables	Number of samples
Original dataset	799	18496
After removal of variables concerning energies other than electricity	682	18496
After removal of imputation and calibration indicator variables	275	18496
After encoding categorical variables	275	18496
After missing Values Management	275	18206
After reduction of target variables	277	18206
After reducing the number of variables	208	18206
After reducing the number of samples	208	17312
End of preprocessing steps	**208**	**17312**

4.3 Variable selection

Variable selection is a crucial step in data preprocessing, aiming to improve the quality of predictive models by identifying the most relevant variables. In our study, we applied several methods to refine our choice of variables.

Correlation across the dataset The first method of variable selection is based on the analysis of the correlation between all the variables in the dataset and the target variables. We first calculated the correlation between each variable and the target variable KWH to identify those with a significant relationship with electricity consumption. The variables whose correlation with KWH was higher than the average of the calculated correlations were selected. Similarly, we evaluated the correlation between each variable and the second target variable, DOLLAREL, to identify the variables that contribute the most to the cost of electricity consumption. The variables whose correlation with DOLLAREL was also higher than the average of the calculated correlations were retained. Finally, the relevant variables for each of the two targets were crossed to obtain their intersection, which made it possible to select 76 important variables for the two target variables, guaranteeing their relevance in both analysis contexts.

Correlation on categorized data The second method of variable selection is based on the categorization of the variables in the dataset and the analysis of the correlation within these categories. This approach allows to structure the selection process by treating the variables according to their specific categories.

The EIA partitioned the variables into 14 distinct categories. We added an additional category specifically dedicated to behavioral variables, comprising 88 manually identified variables. The 15 categories are as follows:

- Administration: This category groups administrative and contextual variables that provide essential information for the identification, climatic classification and urban type of respondents.
- Geography: This category groups geographic variables that provide essential information on the location and geographical characteristics of respondents such as the state of residence.
- Weather: This category groups meteorological variables that provide essential information on the climatic conditions and meteorological characteristics of the areas inhabited by respondents such as outdoor, soil or light bulb temperatures.
- Home characteristics: This category groups variables that provide detailed information on the physical and structural characteristics of respondents' dwellings such as the size and type of dwelling.
- Household characteristics: This category groups together variables that provide detailed information on the socio-demographic and economic characteristics of respondents' households, such as the number of inhabitants, their level of education or household income.

- Appliances: This category groups together variables that provide detailed information on the ownership, use and characteristics of household appliances present in respondents' homes.
- Electronics: This category groups together variables that provide detailed information on the ownership, use and characteristics of electronic devices present in respondents' homes.
- Lighting: This category groups together variables that provide detailed information on the types, use and characteristics of lighting systems present in respondents' homes.
- Air conditioning: This category groups together variables that provide detailed information on air conditioning systems, including their presence, use and characteristics in respondents' homes.
- Water heating: This category groups variables that provide detailed information on water heating systems, including their type, use and characteristics in respondents' homes.
- Space Heating: This category groups together variables that provide detailed information on heating systems, including their type, use and characteristics in respondents' homes.
- Thermostat: This category groups variables that provide information on temperature control devices, including their type, use and characteristics in respondents' homes.
- Energy bills: This category groups variables that provide detailed information on the responsibility for paying energy bills, the presence of a smart meter or the location of the electric vehicle charging.
- Energy assistance: This category groups together variables that provide information on aid and subsidies received for energy consumption.
- Behavioural characteristics: This category groups together variables from all other categories that provide information on households' energy consumption habits and behaviours, including daily routines or usage preferences.

For each category, we calculated the correlation between the variables and the two target variables, then selected those whose correlation was higher than the average within each category. This approach allowed us to retain a total of 137 variables. The table 2 gives a summary of the number of features selected by category.

Information Gain on the entire data set The DecisionTreeRegressor is a regression algorithm that builds a decision tree to predict continuous values, such as energy consumption (KWH) or cost (DOLLAREL). The algorithm assesses the importance of each feature by measuring the reduction in impurity, quantified by the squared error, brought by each division of the tree. At each node, it selects the feature that minimizes this impurity, thus optimizing the predictions. After determining the importance of each feature, we rank these features based on their contribution to reducing the overall squared error. A cumulative curve is then plotted to visualize the contribution of each feature.

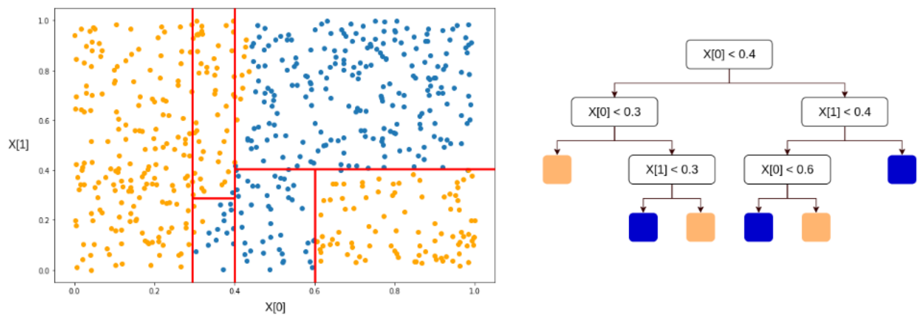


Fig. 3. Illustration of how DecisionTreeRegressor works

Information Gain on categorized data As mentioned earlier, the original data are organized into different categories. For each category, we applied the elbow method to identify the optimal number of variables to keep. This approach consists of plotting the cumulative impurity reduction as a function of the number of retained variables and determining the threshold beyond which the addition of additional variables only marginally improves the impurity reduction. By applying this method individually to each category, we were able to select 83 relevant variables. This approach ensures that the chosen variables contribute significantly to the prediction while maintaining the simplicity and efficiency of the model. The table 2 gives a summary of the number of features selected by category.

Intersection of the 4 methods The intersection of the results obtained by the four selection methods allowed us to identify 17 variables among the initial 206 with the greatest impact on energy consumption. This joint selection highlights the most significant variables by combining the perspectives of the different methods. Among these 17 variables, 9 come from the 65 behavioral variables, highlighting their particular importance in the analysis of energy consumption.

4.4 Descriptive analysis of selected variables

After preprocessing, we selected 17 features with the highest influence on the target attributes. Since the data is unlabeled, we use unsupervised clustering methods to analyze this data.

Determining the optimal number of clusters To determine the optimal number of clusters in our analysis, we used four complementary methods: the elbow method with inertia, the silhouette score, the Davies-Bouldin score, and the Calinski-Harabasz criterion. Each method provides a different perspective

Table 2. Table of results of the two methods using the categories

	Correlation on categorized data		Information Gain on categorized data		
Categories	Number of features	Percentage	Number of features	Percentage	# of features
Behavioral Characteristics	22	33.85%	24	36.92%	65
Appliances	19	43.18%	16	36.36%	44
Electronics	14	37.84%	11	29.73%	37
Home Characteristics	12	46.15%	10	38.46%	26
Air Conditioning	7	38.89%	5	27.78%	18
Space Heating	6	40%	6	40%	15
Household Characteristics	4	33.33%	4	33.33%	12
Energy Bills	2	20%	2	20%	10
Administration	0	0%	2	100%	2
Geography	0	0%	1	33.33%	3
Weather	0	0%	2	40%	5
Water Heating	0	0%	3	60%	5
Lighting	0	0%	4	57.14%	7
Thermostat	0	0%	6	66.66%	9
Energy Assistance	0	0%	6	46.15%	13
Total	**71**	**34.47%**	**81**	**39.32%**	**206**

on the quality of clustering and helps identify the most appropriate number of clusters to best represent the structure of the data.

Elbow Method with Inertia (EM): The elbow method is a classic approach to determining the optimal number of clusters. It relies on inertia, also called the intra-cluster sum of squares. Inertia measures the spread of data points within each cluster: the lower the inertia, the closer the points are to the center of the cluster. Plotting inertia against the number of clusters yields a curve that typically shows a rapid reduction at first, followed by a leveling off. The "elbow" in this curve, where the reduction in inertia begins to decrease significantly, indicates the optimal number of clusters. This point represents a balance between cluster compactness and model complexity.

Silhouette score (SS): The silhouette score evaluates the quality of clustering by measuring the similarity of points within their own cluster relative to the similarity with points in other clusters. The score ranges from -1 to 1: a score close to 1 indicates that points are well clustered in their own cluster and far from other clusters, while a score close to -1 suggests that points might be misclassified. For different numbers of clusters, we calculated the average silhouette score. The number of clusters that maximizes this score is considered optimal, as it indicates the best separation and cohesion between clusters.

Davies-Bouldin Score (DBS): The Davies-Bouldin score is another measure of clustering quality, which assesses the compactness and separation of clusters. It is calculated by measuring the average of the similarity ratios between each cluster and the most similar cluster. A lower score indicates better separation between clusters, suggesting that the clusters are well-distinct and compact. Comparing the scores for different numbers of clusters, the number of clusters

Table 3. Table listing selected variables

Number	Features	Categories	Description
1	BEDROOMS	Home Characteristics	Number of bedrooms
2	COMBODVR	Electronics	Number of cable or satellite boxes with DVR used
3	DRYRUSE	Appliances / Behaviors	Frequency of dryer use per week
4	DWASHUSE	Appliances / Behaviors	Frequency of dishwasher use per week
5	LGTIN1TO4	Lighting / Behaviors	Number of inside light bulbs turned on 1 to 4 hours per day
6	LGTIN4TO8	Lighting / Behaviors	Number of inside light bulbs turned on 4 to 8 hours per day
7	LGTINMORE8	Lighting / Behaviors	Number of inside light bulbs turned on more than 8 hours per day
8	MONEYPY	Household Characteristics	Annual gross household income for the past year
9	MONPOOL	Home Characteristics / Behaviors	Months swimming pool used in the past year
10	NHSLDMEM	Household Characteristics	Number of household members
11	SIZFREEZ	Appliances	Size of most-used freezer
12	TOTROOMS	Home Characteristics	Total number of rooms in the housing unit, excluding bathrooms
13	TOTSQFT_EN	Home Characteristics	Total energy-consuming area (square footage) of the housing unit
14	TVCOLOR	Electronics	Number of televisions used
15	TVONWD2	Electronics / Behaviors	Second-most-used TV usage on weekdays in hours per day
16	USECFAN	Air Conditioning / Behaviors	Most-used ceiling fan usage
17	WASHLOAD	Appliances / Behaviors	Frequency of clothes washer use per week

that minimizes the Davies-Bouldin score is chosen as optimal, as it reflects the best separation of the groups.

Calinski-Harabasz criterion (CHC): The Calinski-Harabasz criterion, also known as the "variance ratio criterion", evaluates the quality of clustering by comparing the intra-cluster variance to the inter-cluster variance. The higher this ratio, the better the separation between clusters and the more compact the clusters are. We calculated this criterion for different numbers of clusters and selected the number that maximizes the Calinski-Harabasz criterion. This indicates that the clusters are well-defined with clear separation and high internal cohesion.

The application of the four methods for selecting the optimal number of clusters led to a convergent conclusion: all methods suggest that the optimal number of clusters is 2. The agreement between these different methods reinforces the conclusion that two clusters are the most appropriate choice to capture the underlying structure of the data. This multi-method consensus validates the robustness and reliability of this solution, thus ensuring an adequate representation of the groups within our analysis.

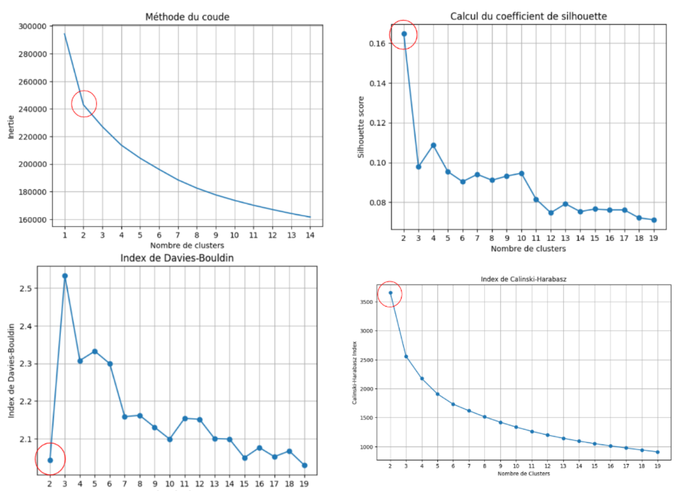


Fig. 4. Graphs of the 4 methods to determine the optimal number of clusters (EM, SS, DBS and CHC)

K-Means K-Means is an unsupervised clustering method that divides a dataset into a fixed number of groups, or clusters, with the goal of minimizing variance within clusters and maximizing variance between them. This algorithm was selected for our study due to its effectiveness in identifying distinct patterns within large datasets. Its simplicity and scalability make it particularly suitable for segmenting households based on energy consumption behaviors, especially when the number of clusters is known or can be estimated. Given our objective of distinguishing between different consumption profiles, K-Means offers a robust method to capture these variations. K-Means with an optimal number of two clusters as determined by preliminary analysis, was chosen to provide a clear and interpretable segmentation of the data.

Kernel K-Means To further refine our analysis, Kernel K-Means was employed to explore the possibility of uncovering more complex, non-linear relationships within the data that might not be detected by the standard K-Means algorithm. Since household energy consumption behaviors can be influenced by intricate and non-linear factors, Kernel K-Means serves as a valuable tool for capturing these subtleties. After initially segmenting the data with K-Means, Kernel K-Means was applied to validate the robustness of the clusters and to investigate any additional, more complex structures. The results confirmed the stability of the identified clusters and provided deeper insights into the patterns of energy consumption.

DBSCAN DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was chosen for this study due to its ability to identify clusters of arbitrary shapes and its robustness in handling noise, which is particularly useful in datasets with varying densities. Unlike K-Means, DBSCAN does not require the number of clusters to be predefined, making it well-suited for exploratory analysis where the cluster structure is unknown. This algorithm is especially advantageous for identifying outliers and isolating them from the main clusters, which is crucial in understanding deviations in household energy consumption behaviors.

In our dataset, applying DBSCAN with an ϵ parameter of 2 and a minimum number of points (minPts) set to 19 revealed four distinct clusters. However, it also classified a significant number of data points as anomalies, indicating the algorithm's strict criteria for forming clusters based on local density. While this feature effectively isolates outliers, it also suggests that further tuning of the ϵ and minPts parameters may be necessary to capture more comprehensive cluster structures. Despite this, DBSCAN provided valuable insights into the presence of small, dense clusters and the identification of outliers, offering an additional perspective on the dataset alongside the results obtained from K-Means and Kernel K-Means.

Cluster analysis Both clustering methods, K-means and Kernel K-means, segmented households into two distinct groups, each highlighting specific energy consumption characteristics.

Households in the first cluster are characterized by smaller dwellings, lower income, and lower ownership and use of amenities. The main drivers of energy consumption in this cluster are related to the essential needs of a dwelling, mainly the use of household appliances for daily tasks such as cooking, refrigeration, and washing. Households in the second cluster, on the other hand, are distinguished by larger dwellings, higher income, and much more intensive use of household appliances and electronics. In this cluster, the main drivers of energy consumption include leisure-related amenities, such as swimming pools, which are largely responsible for the high electricity consumption.

K-Means identified two distinct clusters. The first cluster includes households with an average income of $30,000 to $35,000 per year, an average home size of 1,300 square feet, and consumption habits that are primarily focused on basic needs. The second cluster, on the other hand, includes a wide range of high-income households, with an average annual income of $60,000 to $75,000, an average home size of 2,500 square feet, daily use of appliances such as clothes dryers, and twice the consumption of lighting as the first cluster. The findings thus show a clear distinction between low-income households with basic needs and high-income households with leisure amenities. Kernel K-Means identified a specific subgroup of very wealthy households in the second cluster, thus reducing the total number of samples in this cluster compared to K-Means. This subgroup is distinguished by households with even larger dwellings, with an average area of 260 square feet, and a higher use of all amenities. This means that Kernel K-Means was able to identify a group of extremely wealthy households, whose energy consumption habits are even more marked by amenities and luxurious consumption behaviors.

The differences between the two clusters highlight the various aspects of household energy consumption. For the first cluster, energy consumption is mainly due to essential needs, reflecting households living in modest housing with lower incomes. Strategies to reduce energy consumption in this cluster could focus on the energy efficiency of household appliances and daily practices. For the second cluster, high energy consumption is mainly attributed to leisure facilities such as swimming pools or satellite boxes, characteristic of high-income households living in large housing. Efforts to reduce energy consumption in this group could include measures to make these facilities more efficient or promote less energy-intensive alternatives.

To complement this analysis, we also applied the DBSCAN algorithm, which also identified two distinct clusters. However, it is worth noting that the majority of the data points were classified as anomalies. Unlike the results of previous algorithms, the two clusters detected by DBSCAN mainly group households with lower incomes. Although these households have differences in terms of income and housing size, they show similar behaviors in terms of low use of household equipment and appliances. A notable distinction is that one of these clusters does not own a television at all.

In summary, this analysis reveals that energy consumption behaviors differ greatly by income and housing size, directly influencing the potential energy-saving strategies to be adopted for each group. These results highlight the importance of targeting specific energy efficiency measures adapted to the characteristics and needs of different segments of the population, in order to maximize energy savings and improve environmental sustainability.

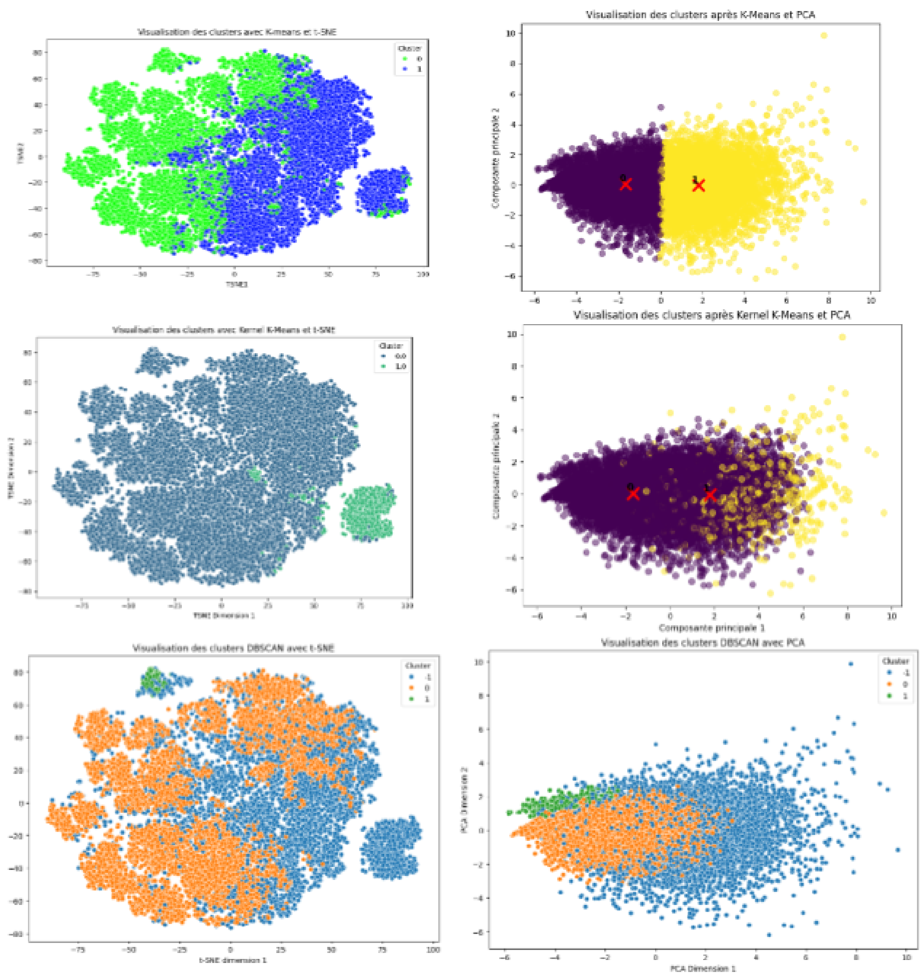


Fig. 5. Visualizations of clustering with K-Means, Kernel K-Means and DBSCAN

5 Analysis of behavioral variables

Now, we focused our attention on behavioral variables among those that were selected. Recall that the initial objective of our study was to analyze the influence of behavior on energy consumption. Among the 17 variables resulting from

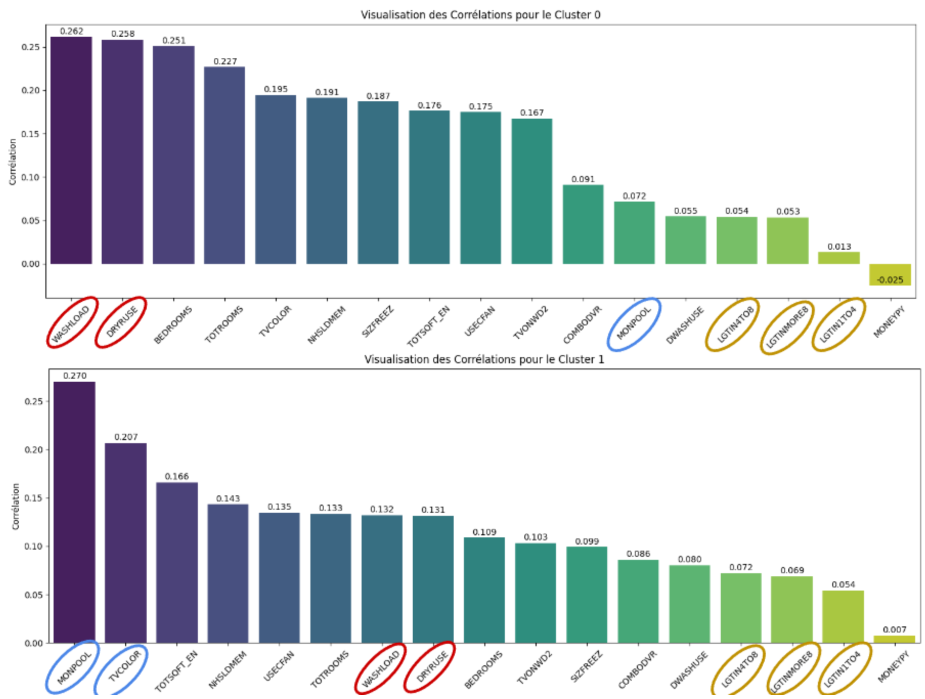


Fig. 6. Correlations of the clustering of the selected variables with K-Means

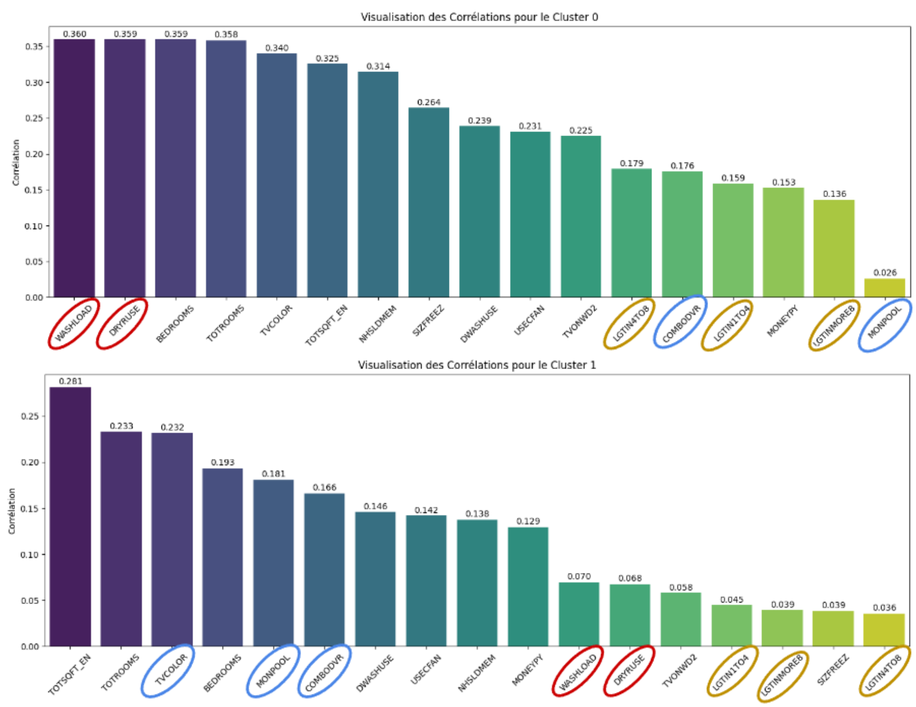


Fig. 7. Correlations of the clustering of the selected variables with Kernel K-Means

the selection of variables, 9 are specifically related to household behaviors. By isolating these behavioral variables, we can examine in more detail how household habits and practices impact their energy consumption. This step is crucial to understand the specific behavioral factors that contribute to the variation in energy consumption between different households.

Table 4. Table listing selected behavioral variables

Number	Features	Description
1	DRYRUSE	Frequency of dryer use per week
2	DWASHUSE	Frequency of dishwasher use per week
3	LGTIN1TO4	Number of inside light bulbs turned on 1 to 4 hours per day
4	LGTIN4TO8	Number of inside light bulbs turned on 4 to 8 hours per day
5	LGTINMORE8	Number of inside light bulbs turned on more than 8 hours per day
6	MONPOOL	Months swimming pool used in the past year
7	TVONWD2	Second-most-used TV usage on weekdays in hours per day
8	USECFAN	Most-used ceiling fan usage
9	WASHLOAD	Frequency of clothes washer use per week

5.1 Determining the optimal number of clusters

As before, we determined the optimal number of clusters using several evaluation methods. The elbow method suggested that the optimal number of clusters is 6, while the silhouette score indicated 3 clusters. In contrast, the Davies-Bouldin index recommended 10 clusters, and the Calinski-Harabasz criterion pointed to 2 clusters. These discrepancies highlight the importance of considering multiple criteria for a robust assessment of the number of clusters, with each method having its own criteria and interpretations of the data.

5.2 Two clusters

The analysis revealed two main clusters, as previously observed. The results obtained with the K-means and Kernel K-means methods show a similarity with the clusters identified in previous analyses. Both methods confirmed the separation between low- and high-income households. Kernel K-means once again highlights a specific subgroup among the wealthiest households.

5.3 Three clusters

The analysis with three clusters reveals notable distinctions in household energy consumption behaviors. This clustering highlights the predominant impact of swimming pool use on energy consumption.

The most energy-intensive cluster is characterized by high pool use, making it the main factor influencing energy consumption in this group. Even in clusters

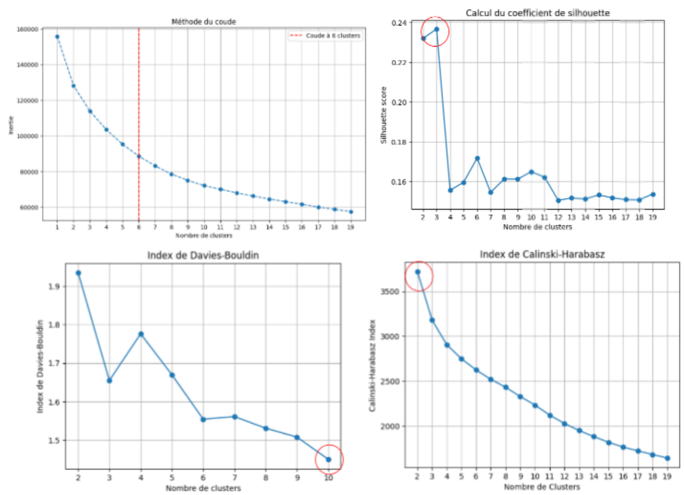


Fig. 8. Graphs of the 4 methods for determining the optimal number of clusters for behavioral variables

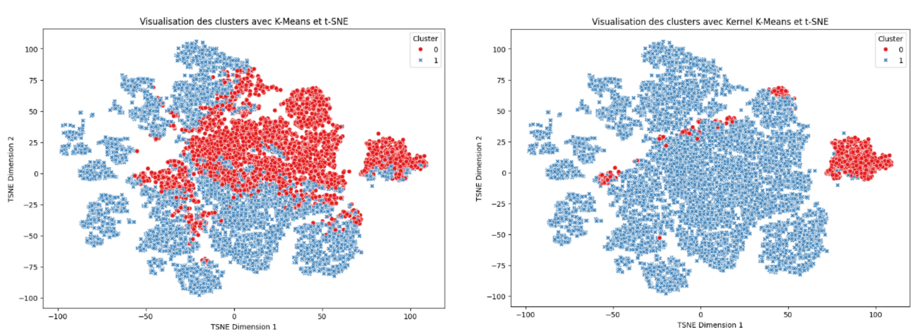


Fig. 9. Visualization of the clustering of behavioral variables into 2 clusters

where pool use is less frequent, this variable remains strongly correlated with energy consumption. This observation indicates that, regardless of the level of pool use, it exerts a significant influence on household energy behaviors, highlighting the importance of this equipment in energy consumption analyses.

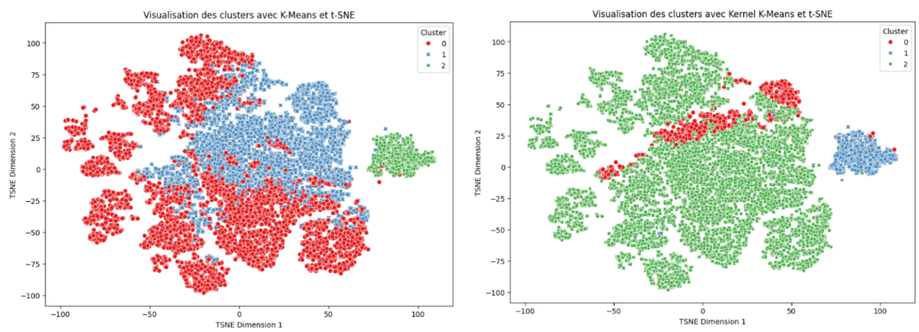


Fig. 10. Visualization of the clustering of behavioral variables into 3 clusters

5.4 Six clusters

By applying K-means and Kernel K-means algorithms to segment the data into 6 clusters, we further analyzed household energy consumption behaviors. The results show that both clustering methods produce similar segmentation profiles.

Swimming pool use is confirmed as the most important factor for energy consumption, exerting a major influence on overall consumption in all clusters. In particular, wealthy households, a significant proportion of which reside in Florida, tend to use their pools almost half of the year, leading to high energy consumption. These households often have large homes, which could also explain their high use of lights, or they are simply less sensitive to the energy impact of these devices. Families with children are characterized by a high use of kitchen and household appliances, using machines such as washers and dryers up to almost twice a day for each appliance. They also have several televisions, one of which is often used for children's video games, which contributes to high energy consumption. For more modest households, household appliances are the main consumption factor, followed by television. Families living in small homes show energy behaviors where the use of household appliances and television is predominant. For retirees, television is the main energy consumption factor, with use exceeding 10 hours per day. This behavior reflects their tendency to spend a large part of the day in front of the screen.

This analysis thus reveals significant variations in energy behaviors according to resources, household priorities, and geographical contexts.

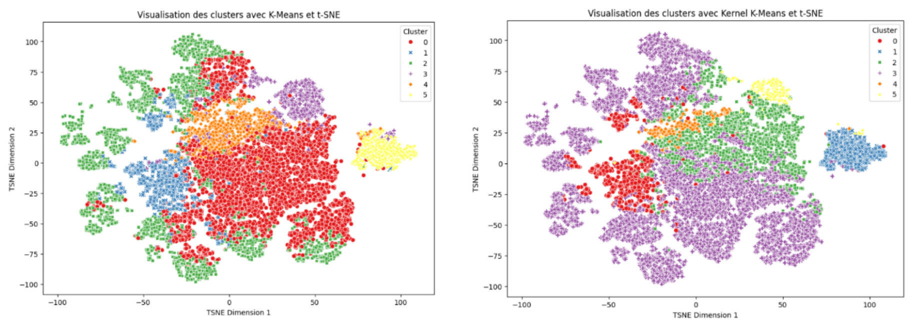


Fig. 11. Visualization of the clustering of behavioral variables into 6 clusters

5.5 Ten clusters

By applying K-means and Kernel K-means algorithms to segment the data into 10 clusters, we were able to explore the subtleties of household energy behaviors in more depth. Both methods highlighted several significant distinctions, while confirming some trends observed in previous analyses.

Swimming pool use remains the most energy-intensive factor, clearly standing out in all clusters. However, other energy behaviors are more finely distributed in this 10-cluster clustering.

The notable distinction concerns the use of lights. This clustering differentiates households based on the duration of light use. For example, intensive use of lights for a short period of time (less than 8 hours per day) in cluster 7 was found to be more energy-intensive than prolonged use of lights for a longer period of time (more than 8 hours per day) in cluster 9. With further investigation, this can be explained, in the case of cluster 7, by the fact that 37.55% of households do not use LED bulbs at all, compared to only 9% in cluster 9. Furthermore, 72.27% of households in cluster 9 use LED bulbs predominantly, compared to only 15.92% in cluster 7. The duration of use of lights may therefore not be the determining factor but rather the type of bulbs.

Finally, television, as in previous analyses, is here relegated to the bottom of the ranking of energy behaviors, partly due to the widespread adoption of energy-efficient televisions, such as LED televisions. However, about 10% of households still own energy-intensive televisions, such as plasma or CRT televisions, which have a greater impact on overall energy consumption compared to other household appliances and uses.

By segmenting the data into 10 clusters using the K-means and Kernel K-means algorithms, the analysis revealed subtle distinctions in household energy behaviors. This segmentation notably highlighted marked variations in the use of lights, where intensive use over a short period with non-LED bulbs proved to be more energy-intensive than prolonged use with LED bulbs.

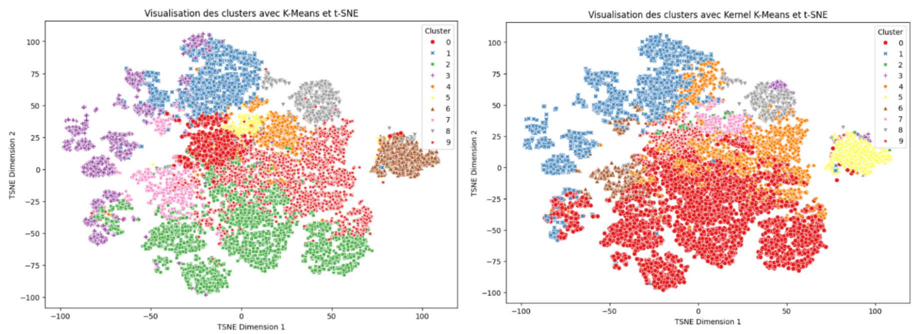


Fig. 12. Visualization of the clustering of behavioral variables into 10 clusters

5.6 DBSCAN

Clustering performed with the DBSCAN algorithm, using the parameters eps=1 and min_samples=18, allowed defining two distinct clusters of households without swimming pools, one of which is also characterized by the absence of ceiling fans. This segmentation highlights groups of households with particular energy behaviors, particularly in the absence of equipment that usually consumes energy.

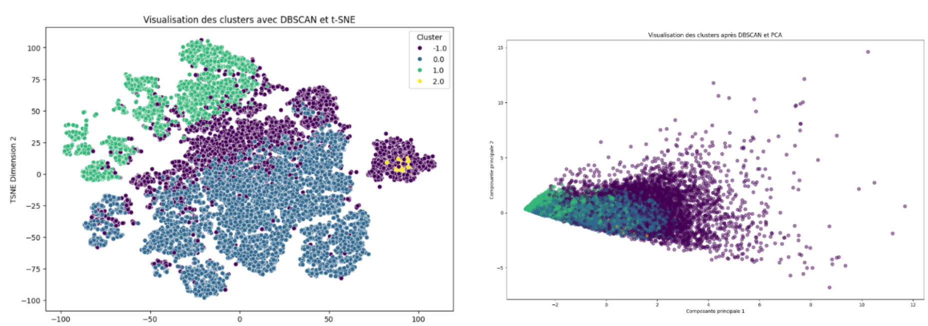


Fig. 13. Visualization of behavioral variables clustering with DBSCAN

6 Discussion and recommendations

6.1 Discussion

In this study, we identified several key behaviors influencing residential energy consumption. This section highlights the main findings in relation to relevant data and trends.

Pool usage The analysis revealed that swimming pool usage is the most energy-intensive factor in residential energy consumption. Even in clusters where pool use is less frequent, this variable remains strongly correlated with energy consumption. This indicates that, regardless of how frequently pools are used, they have a significant impact on household energy behaviors. Moreover, high pool usage is more energy-consuming than the intensive use of any other variable analyzed, highlighting the crucial role of swimming pools in energy consumption analyses. This observation is reinforced by market data [19], which shows a 101% increase in the sale of heat pumps for swimming pools between 2019 and 2021. In the United States, the residential swimming pool market is significant, with 10.7 million swimming pools, of which 10.4 million are residential, and almost 59% are in-ground. Indeed, the United States is the country with the most private swimming pools in the world. Wealthy households, many of which are located in Florida, tend to use their pools for almost half of the year, leading to high energy consumption. These households often have larger homes, which may explain their higher use of both pools and lighting. Alternatively, they may be less sensitive to the energy impact of these amenities.

Appliances Families with children exhibit high energy consumption due to their extensive use of kitchen and household appliances. These families often operate washing machines and dryers nearly twice daily for each appliance. In contrast, other types of households tend to have more standard usage patterns for these appliances. This heightened use of household devices significantly contributes to their overall energy consumption, illustrating how family structure can influence energy behaviors.

Lighting A notable distinction concerns the use of lights. The analysis shows that the energy consumption associated with lighting is less about the number of hours lights are used and more about the type of bulbs employed. Approximately 37% of households use few or no LED bulbs during the day. Additionally, around 14% of individuals who leave lights on throughout the night do not use LED bulbs, which is significant given the high electricity consumption of incandescent and halogen bulbs.

Television Although the impact of television on energy consumption may be relatively small when used moderately, its cumulative effect can become significant when usage is widespread. This is particularly evident among retirees, who are the largest consumers of television, with usage exceeding 10 hours per day. This behavior reflects their tendency to spend a substantial part of the day in front of the screen. Families with children also contribute to high energy consumption through television use. They often own multiple TVs, with one frequently used for children's video games, further increasing energy consumption.

Kitchen Although kitchen appliances are not specifically covered in this study, they still play a role in household energy consumption. The relatively small impact of cooking on overall electricity consumption in American households is largely due to their cooking habits. On average, Americans spend about 1 hour at the table, compared to over 2 hours for the French [41]. As noted, "Americans have very nice kitchens, but they don't cook" [42]. The preference for junk food and fast food, especially in low-income families, limits the use of kitchen appliances, as fresh produce is often more expensive and less accessible [40]. However, families with children typically use kitchen appliances more frequently, which increases their overall energy consumption in this area.

Smart appliances Although smart devices were not widely featured in this study due to their low individual power consumption and low usage in U.S. households, they are present in the original dataset, notably for controlling temperature, television, lights and security. 62% of the households studied do not have connected devices, which significantly limits their ability to monitor and optimize their energy consumption. Furthermore, although 26% of the sample have a connected meter, only 8% check their consumption at regular intervals. These relatively low figures highlight an under-exploited potential.

The image below 14 shows that even if some households make frequent use of the swimming pool, washing machine, lights or television, they are not necessarily the ones that consume the most electricity. We could deduce that electricity consumption is more influenced by the quality of the equipment than by its intensive use.

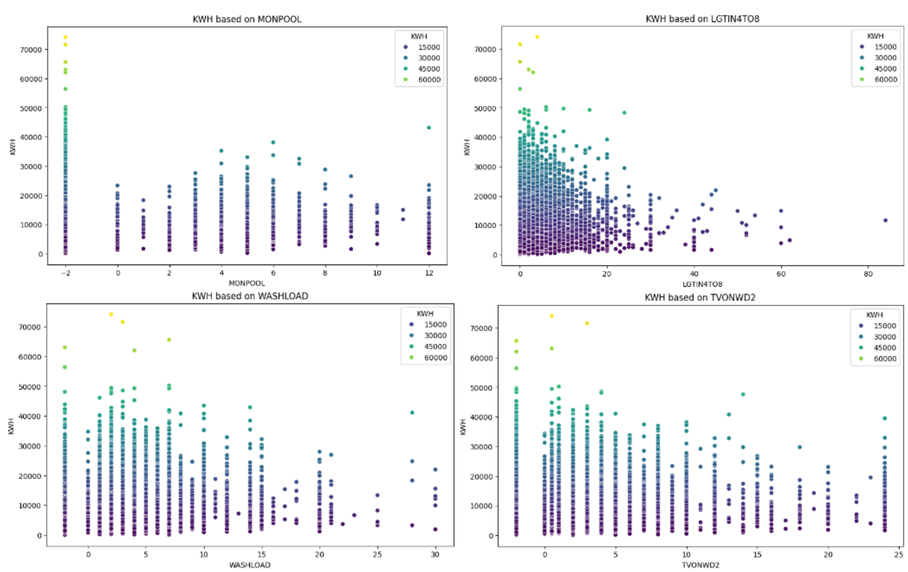


Fig. 14. Distribution of some variables as a function of electrical consumption

6.2 Recommendations

Based on the findings of this study, several recommendations emerge to enhance energy efficiency and reduce consumption.

For individuals, a multifaceted approach can significantly improve energy use and contribute to a healthier lifestyle. Promoting outdoor physical activities reduces reliance on energy-intensive devices while positively impacting overall well-being. Limiting screen time and exploring alternative leisure activities can further decrease energy consumption. Incorporating smart technology, such as programmable thermostats and automated lighting controls, allows for more efficient management of energy use by ensuring devices operate only when necessary. Additionally, adopting energy-efficient practices like using LED bulbs, which use up to 10 times less energy than incandescent bulbs and 6–8 times less than halogen bulbs [23], can greatly minimize waste. Prolonged lighting, including outdoor lights left on at night, accounts for 10–15% of average residential electricity consumption [22], so switching to LED bulbs and implementing automated lighting controls can result in significant savings. For households with swimming pools, strategies such as optimizing heating systems or using thermal blankets can notably reduce both energy consumption and water evaporation. These measures can help in reducing overall water consumption as well [20]. In the kitchen, adopting more energy-efficient cooking practices, such as using convection ovens or induction cook-tops, and optimizing their efficiency, can further reduce energy consumption. Household appliances also play a significant role in overall energy use. Modern, high-efficiency appliances, certified by labels like Energy Star, can cut energy consumption by up to 15% compared to older models [21]. Therefore, investing in these appliances and utilizing their energy-saving modes, such as eco-mode on dishwashers and washing machines, can contribute to substantial long-term savings. By making these adjustments, individuals can play a crucial role in reducing household energy consumption while benefiting from improved well-being.

For authorities, fostering energy efficiency and sustainability involves a multifaceted approach. One key strategy is to focus on public infrastructure. Developing public swimming facilities, rather than encouraging private pools, can reduce both water and energy consumption associated with individual household pools. Additionally, increasing the purchase price of swimming pools could potentially decrease the number of new pool installations, leading to further reductions in both energy and water consumption, provided this approach is politically acceptable. Improving home insulation by enhancing thermal efficiency in walls, windows, and roofs can substantially decrease heating and cooling needs, leading to significant energy savings. Promoting renewable energy is another critical area. Investing in solar panels for homes can provide a sustainable energy source and lessen dependence on non-renewable resources. Supporting energy-efficient building standards for new constructions and renovations—such as promoting passive or low-energy buildings—can further contribute to reduced energy consumption and long-term sustainability. Encouraging the adoption of energy-efficient appliances through incentives like rebates or subsidies can drive broader use of technologies that significantly cut energy usage. Public awareness and education play a crucial role as well. Implementing educational campaigns and workshops on energy conservation and smart technology benefits can empower individuals to adopt energy-saving practices. Promoting alternatives to television, such as social or physical activities, could not only reduce energy consumption but also provide benefits for health and well-being. Encouraging these behaviors can thus contribute to energy savings while improving the quality of life for retirees. Developing infrastructure for electric vehicle charging and investing in expanding public transportation networks can reduce the energy consumption linked to personal vehicles. By integrating these strategies, authorities can create a supportive environment that promotes both individual and communal efforts toward greater energy efficiency and sustainability, ultimately benefiting both the economy and the planet.

7 Conclusion

This study conducted an in-depth analysis of residential energy consumption using preprocessing, feature selection and clustering methods to understand the influence of household behaviors and identify distinct energy consumption groups.

We began with rigorous data preprocessing, reducing the number of samples from 18,496 to 17,312 and the variables from 799 to 310. Four feature selection methods reduced the number of relevant variables to 28, 12 of which were behavioral. The application of clustering algorithms, such as K-Means and Kernel K-Means, revealed significant variations in household energy behaviors. The analyses identified several key behaviors based on their energy impact. Among these behaviors, the use of swimming pools stands out as the most energy-consuming factor. Household appliances also have a significant impact, while lighting, especially when using non-LED bulbs, contributes significantly to energy consumption. The use of kitchen appliances and watching television, although less energy-consuming, also contribute to overall consumption. This distribution highlights the importance of implementing appropriate strategies to manage energy consumption efficiently. Leisure facilities, especially swimming pools, require targeted interventions to improve their energy efficiency. Household appliances should be optimised to minimise their impact, while measures such as adopting LED bulbs and implementing automated control systems can offer significant savings in lighting.

In the future, digital technologies, such as video streaming and cloud computing, which are highly resource-intensive, could significantly increase residential energy consumption. In 2023, the IT sector accounts for about 10% of global electricity consumption, up from 7% in 2016, and this share could reach 20% by 2030. Video streaming, which constitutes about 80% of global web traffic, is a major contributor to this trend [43]. The increasing impact of these technologies on electricity consumption requires special attention to develop effective strategies to mitigate their effect on energy demand.

On another side, it is relevant to note that electricity consumption prediction could benefit from the integration of emerging digital technologies. These technologies play an increasingly crucial role in managing and reducing residential energy consumption. The Internet of Things (IoT) and connected objects enable real-time data collection on energy consumption, providing opportunities for more detailed analyses and more accurate forecasts. Home energy management systems (HEMS) can monitor and optimize energy usage, while smart grids enable a more balanced and efficient distribution of electricity. The Internet of Behavior (IoB) could also play a crucial role by enabling a deep understanding of user habits and behaviors. Unfortunately, for this research, we were unable to access such detailed data, which limits our ability to take full advantage of these emerging technologies. However, they could transform the way we anticipate energy needs and optimize consumption management, offering powerful tools to improve energy efficiency and reduce household carbon footprints in the future.

Today, as we reach Earth Overshoot Day in 2024, this study highlights the critical environmental impact of our energy behaviors. Earth Overshoot Day marks the date when humanity has consumed all the resources that Earth can renew in one year. Adopting targeted strategies to optimize energy consumption at the household level is essential to foster greater awareness and more sustainable practices, helping to reduce our environmental footprint and preserve the planet's resources.

References

1. U.S. Energy Information Administration, 2020 Residential Energy Consumption Survey, `https://www.eia.gov/consumption/residential/data/2020/`, last accessed 01/08/2024
2. Institut National d'Etudes Démographiques, 2024 : les Nations unies publient de nouvelles projections de population mondiale, `https://www.ined.fr/fr/tout-savoir-population/memos-demo/focus/2024-les-nations-unies-publient-de-nouvelles-projections-de-population-mondiale`, last accessed 30/08/2024
3. Statista, Evolution du nombre d'habitants sur Terre entre 1950 et 2024 et projections jusqu'en 2100, `https://fr.statista.com/statistiques/564933/population-mondiale-jusqu-en-2080/`, last accessed 01/08/2024
4. Nations Unies, Population, `https://www.un.org/fr/global-issues/population`, last accessed 01/08/2024
5. Pablo-Romero, M. del P., Pozo-Barajas, R., Yñiguez, R., Global changes in residential energy consumption, Energy Policy (2016), `https://doi.org/10.1016/j.enpol.2016.10.032`
6. Chareyron, D., Horsin-Molinaro, H., Multon, B., Concepts et chiffres de l'énergie: la consommation de l'électricité domestique en France, Culture Sciences Physiques, ENS Lyon, `https://culturesciencesphysique.ens-lyon.fr/ressource/chiffres-energie-electricite-domestique.xml`, last accessed 01/08/2024
7. Taconet, C., Objets connectés : 50 milliards d'émetteurs de CO2 ?, Télécom SudParis, `https://www.telecom-sudparis.eu/actualite/objets-connectes-50-milliards-demetteurs-de-co2/`, last accessed 30/08/2024

8. Sun, Jiayi, Gan, Wensheng, Chao, Han-Chieh, Philip, S Yu, Ding, Weiping. Internet of behaviors: A survey. IEEE Internet of Things, vol. 10, no. 13, pp. 11117 – 11134, 2023.
9. Zhao, Q., Li, G., Cai, J., Zhou, M., and Feng, L. A Tutorial on Internet of Behaviors: Concept, Architecture, Technology, Applications, and Challenges. IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 1227 –1260, 2023.
10. Ziani, L., Khanouche, M. E., Belaid A.: Internet of behaviors: A literature review of an emerging technology. In 1st Int. Conf. on Big Data, IoT, Web Intelligence and Applications, pp. 42–47, Sidi Bel Abbes, Algeria, 2022,
11. Javaid, M., Haleem, A., Singh, R.P., Khan, S., Suman, R.: An extensive study on internet of behavior (iob) enabled healthcare-systems: Features, facilitators, and challenges. BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 100085, 2023.
12. Mezair, Tinhinane, Djenouri, Youcef, Belhadi, Asma, Srivastava, Gautam, Lin, Jerry Chun-Wei. Towards an Advanced Deep Learning for the Internet of Behaviors: Application to Connected Vehicles. ACM Transactions on Sensor Networks, vol. 19, no. 2, pp. 1–18, 2022.
13. Song, Qun, Tan, Rui, Wang, Jianping. Towards Efficient Personalized Driver Behavior Modeling with Machine Unlearning. In Proceedings of Cyber-Physical Systems and Internet of Things, pp. 31–36, 2023.
14. Embarak, O.H.: Internet of behaviour (IoB)-based AI models for personalized smart education systems. Procedia Computer Science 203, 103–110, 2022.
15. Embarak, Ossama. An adaptive paradigm for smart education systems in smart cities using the internet of behaviour (IoB) and explainable artificial intelligence (XAI). 8th Int. Conference on Information Technology Trends, pp. 74–79, 2022.
16. Sani, H. M., Tehrani, S. O., Behkamal, B., Amintoosi, H.: Extracting Effective Features for Descriptive Analysis of Household Energy Consumption Using Smart Home Data. In: Grandinetti, L., et al. (eds.) TopHPC 2019, CCIS, vol. 891, pp. 240–252. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33495-6_18
17. Heinrich, M., Ruellan, M., Oukhellou, L., Samé, A., Lévy, J.-P.: From energy behaviours to lifestyles: Contribution of behavioural archetypes to the description of energy consumption patterns in the residential sector. Energy and Buildings, 2022, 269, pp.112249. 10.1016/j.enbuild.2022.112249. hal-03892707f
18. Sanquist, T.F., Orr, H., Shui, B., Bittner, A.C.: Lifestyle factors in U.S. residential electricity consumption. Energy Policy, **42**, 354–364 (2012). https://doi.org/10.1016/j.enpol.2011.11.092
19. Aquark, Analyse du marché: Quelle est la taille du marché mondial de la piscine ?, https://www.aquark.com/fr/analyse-du-marche-mondial-de-la-piscine/, last accessed 01/08/2024
20. Guide Piscine, L'impact d'une couverture sur la consommation en eau et en électricité d'une piscine, https://www.guide-piscine.fr/pro/marche-de-la-piscine/piscine-eco-responsable/l-impact-d-une-couverture-sur-la-consommation-en-eau-et-en-electricite-d-une-piscine-5620_A, last accessed 01/08/2024
21. U.S. Energy Information Administration, A call to action on efficient and smart appliances, https://www.iea.org/articles/a-call-to-action-on-efficient-and-smart-appliances, last accessed 01/08/2024

22. Engie, LED, halogène, connectée… quelle est la consommation d'une ampoule ?, https://particuliers.engie.fr/economies-energie/conseils-economies-energie/conseils-calcul-consommation/consommation-ampoule.html, last accessed 01/08/2024
23. Chesnais, E.: Ampoules LED De belles économies, UFC Que Choisir, https://www.quechoisir.org/guide-d-achat-ampoules-basse-consommation-led-n11547/, last accessed 01/08/2024
24. Labiadh M.: Methodology for construction of adaptive models for the simulation of energy consumption in buildings. Modeling and Simulation. Université de Lyon, 2021. English. NNT : 2021LYSE1158. tel-03662903
25. Elayan, H., Aloqaily, M., Karray, F., Guizani, M.: Decentralized IoB for Influencing IoT-based Systems Behavior. In: 2022 IEEE International Conference on Communications (ICC 2022), IEEE (2024). https://doi.org/10.1109/ICC45855.2022.9839241
26. Elayan, H., Aloqaily, M., Karray, F., Guizani, M.: Internet of Behavior and Explainable AI Systems for Influencing IoT Behavior. IEEE Network (2023). https://doi.org/10.1109/MNET.009.2100500
27. Wang, J., Chen, X., Zhang, F., Chen, F., Xin, Y.: Building Load Forecasting Using Deep Neural Network with Efficient Feature Fusion. Journal of Modern Power Systems and Clean Energy, vol. 9, no. 1 (2021). https://doi.org/10.35833/MPCE.2020.000321
28. Chan, J. W., Yeo, C. K.: Electrical Power Consumption Forecasting with Transformers. In: 2022 IEEE Electrical Power and Energy Conference (EPEC). IEEE (2022). https://doi.org/10.1109/EPEC56903.2022.10000228
29. Hadjout, D., Torres, J. F., Troncoso, A., Sebaa, A., Martínez-Alvarez, F.: Electricity consumption forecasting with outliers handling based on clustering and deep learning with application to the Algerian market. Expert Systems with Applications. Elsevier (2021). https://doi.org/10.1016/j.eswa.2023.120123
30. Hadjout, D., Torres, J. F., Troncoso, A., Sebaa, A., Martínez-Alvarez, F.: Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market. Elsevier (2021). https://doi.org/10.1016/j.energy.2021.123060
31. Syed, D., Abu-Rub, H., Ghrayeb, A., Refaat, S. S.: Household-Level Energy Forecasting in Smart Buildings Using a Novel Hybrid Deep Learning Model. IEEE Access (2021). https://doi.org/10.1109/ACCESS.2021.3061370
32. Alhussein, M., Aurangzeb, K., Haider, S. I.: Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting. IEEE Access (2020). https://doi.org/10.1109/ACCESS.2020.3028281
33. Li, X., Zhong, Y., Shang, W., Zhang, X., Shan, B., Wang, X.: Total electricity consumption forecasting based on Transformer time series models. Procedia Computer Science **214**, 312–320 (2022). https://doi.org/10.1016/j.procs.2022.11.180
34. Liu, Y., Zhang, D., Gooi, H. B.: Optimization Strategy Based on Deep Reinforcement Learning for Home Energy Management. CSEE Journal of Power and Energy Systems **6**(3), 572–581 (2020). https://doi.org/10.17775/CSEEJPES.2019.02890
35. Kim, S.-H., Lee, C., Youn, C.-H.: An Accelerated Edge Cloud System for Energy Data Stream Processing Based on Adaptive Incremental Deep Learning Scheme. IEEE Access **8**, 195341 (2020). https://doi.org/10.1109/ACCESS.2020.3033771
36. Ziani, L., Khanouche, M. E., Belaid, A.: Internet of Behaviors: A literature review of an emerging technology. In: Proceedings of the First International Conference on Big Data, IoT, Web Intelligence and Applications (BIWA), pp. 1–8. IEEE (2022). https://doi.org/10.1109/BIWA57631.2022.10037987

37. Sun, J., Gan, W., Chao, H.-C., Yu, P. S., Ding, W.: Internet of Behaviors: A Survey. IEEE Internet of Things Journal **10**(13), 1–12 (2023). https://doi.org/10.1109/JIOT.2023.3247594
38. Zhao, Q., Li, G., Cai, J., Zhou, M., Feng, L.: A Tutorial on Internet of Behaviors: Concept, Architecture, Technology, Applications, and Challenges. IEEE Communications Surveys & Tutorials **25**(2), 1–22 (2023). https://doi.org/10.1109/COMST.2023.3246993
39. Tutunea, M-F.: Internet of Behavior (IoB) – an alternative for differentiation in the digital environment, Studia UBB Negotia, vol. 67, issue 3 (September) 2022, pp. 7-26, https://doi.org/10.24193/subbnegotia.2022.3.01
40. EuropUSA, Etats-Unis: les habitudes alimentaires, https://www.europusa.com/vivre-aux-etats-unis/vie-personnelle-aux-usa/vie-quotidienne/etats-unis-les-habitudes-alimentaires/, last accessed 01/08/2024
41. SudOuest, Les Français passent deux fois plus de temps à table que les Américains, https://www.sudouest.fr/premium/art-de-vivre/les-francais-passent-deux-fois-plus-de-temps-a-table-que-les-americains-3136883.php, last accessed 01/08/2024
42. Christina Gierse, Les Français et la nourriture aux Etats-Unis: un sujet « touchy », Studyrama, https://www.studyrama.com/pro/destination/les-francais-et-la-nourriture-aux-etats-unis-un-sujet-touchy-21891.html, last accessed 01/08/2024
43. Christophe Magdelaine, Streaming : quelles émissions de CO2 et consommation d'énergie ?, https://www.notre-planete.info/actualites/247-streaming-Internet-electricite-CO2, last accessed 01/08/2024
44. Scikit Learn, IsolationForest, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html, last accessed 30/08/2024