Developing an information and recommendation system for saving and ranking building energy consumption by use of occupant behavior: A Data Mining and MCDM approach

A Summary of the Master’s Thesis

Morteza Emadi

Main Supervising Team:

Prof. Pouria Ahmadi (Uni. of Tehran)

Prof. Mehdi Sadatrasoul (Kharazmi Uni.)

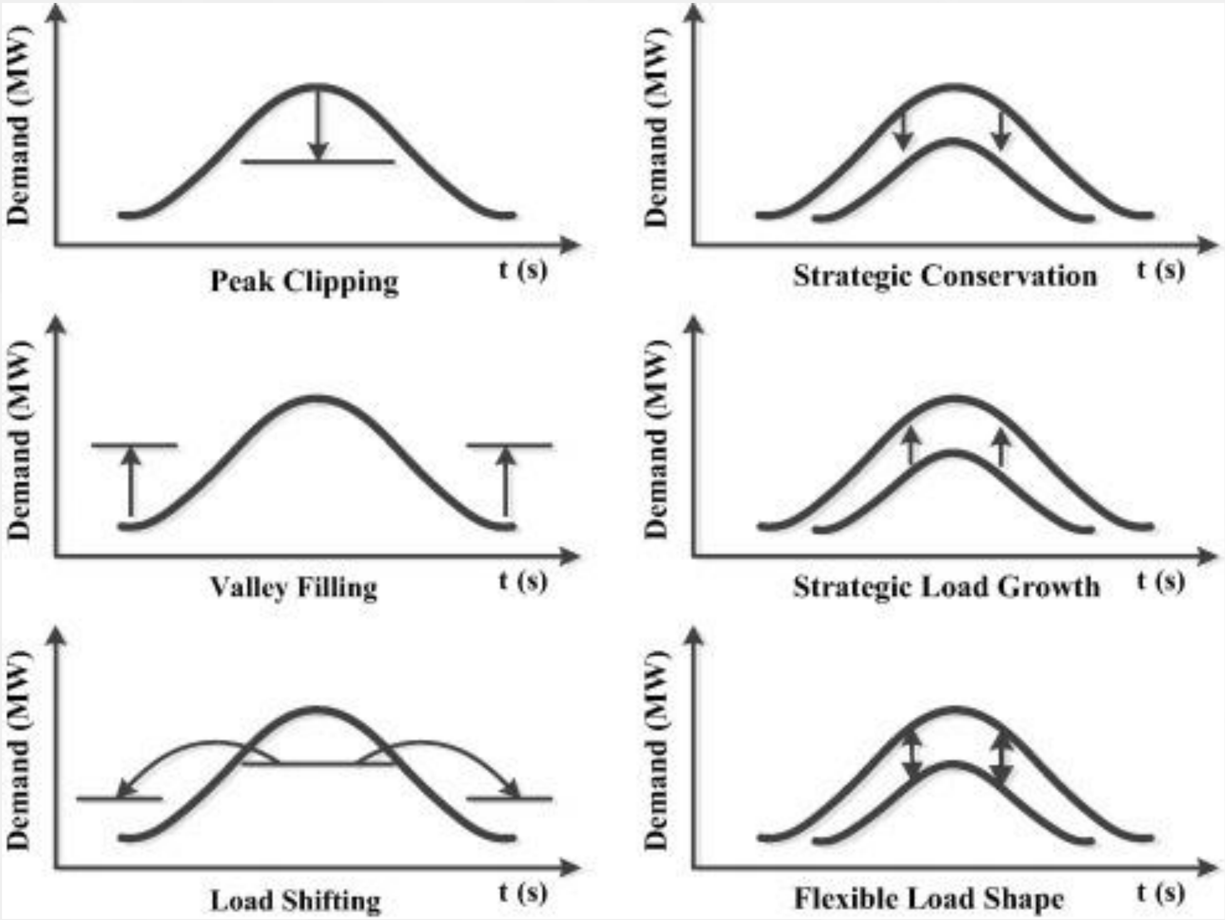
Kharazmi University

Management Faculty, Kharazmi Univeristy

Short Introduction

Demand-side management

It's a set of measures that allow energy consumers to adjust their demand, especially during peak consumption periods, to reduce their overall energy consumption. It involves two primary activities: **demand response programs**, such as load shifting, and **energy conservation**.(Momoh, 2018)



Statement of the Problem

Residential buildings account for a significant portion of global energy consumption, with an estimated 29% of global energy use occurring in residential buildings in 2020 (International Energy Agency, 2021). Addressing the issue of energy consumption in residential buildings has become increasingly crucial due to their significant contribution to environmental degradation and energy insecurity. Despite the various energy-saving solutions available, the potential of demand-side management, particularly with an emphasis on occupant behavior, has not been fully explored. Without a comprehensive understanding of occupants' energy consumption behavior, designing and implementing effective behavior modification strategies becomes challenging when relying on traditional simulation models. Therefore, this study aims to address this research gap by investigating the energy consumption patterns in residential buildings, identifying behavioral patterns through data-driven methods, and subsequently recommending tailored behavior modifications for building occupants. To encourage energy-saving practices, an innovative ranking system will be introduced, fostering competition among buildings.

Smart Measurement and Data-Driven approach

Data-driven modeling that uses machine learning (ML) algorithms can adapt its prediction and reaction with short-term historical data, without obtaining an explicit definition of occupant behavior and events; as such, they have very little dependence on the context.

Saving by behavior modification

The behavior of each occupant is unique by itself. Therefore, the usefulness of the holistic Occupant Behavior classification is questionable i.e., there are no certain classified predefined energy-saving and energy-wasting behavior categories. So, Data Mining particular residents' behavior is a prerequisite for recommending any behavior modification to them.

Objectives of the Study

Primary

1. Discovering behavioral opportunities for reducing energy consumption and finding exact saving potentials by using occupant energy consumption data

2. Examining the possibility of **ranking** and comparing the behavior of different buildings with respect to behavioral opportunities for reducing energy consumption

Secondary

3. Evaluating different data mining methods to **eliminate non-behavioral variables** in energy consumption analysis

4. Examining the possibility of discovering meaningful rules among energy consumers' behaviors

Gaps and Contribution

1) Filling the gaps

Using **Multi-Criteria Decision-Making** methods to rank energy consumption behavior in building **along** with data mining to improve occupants' consumption

2) Filling the gaps

Detecting the on/off status of the radiators based on three thermal data and employing its consumption in consumption behavior analysis

3) Contribution

Qualifying the input data to association rules via a heuristic method that has several thresholds of the temporal resolution type and sensor values

4) Contribution

Using an up-to-date dataset with unique sampling rate which also overcame the big data challenges

Introducing the Dataset

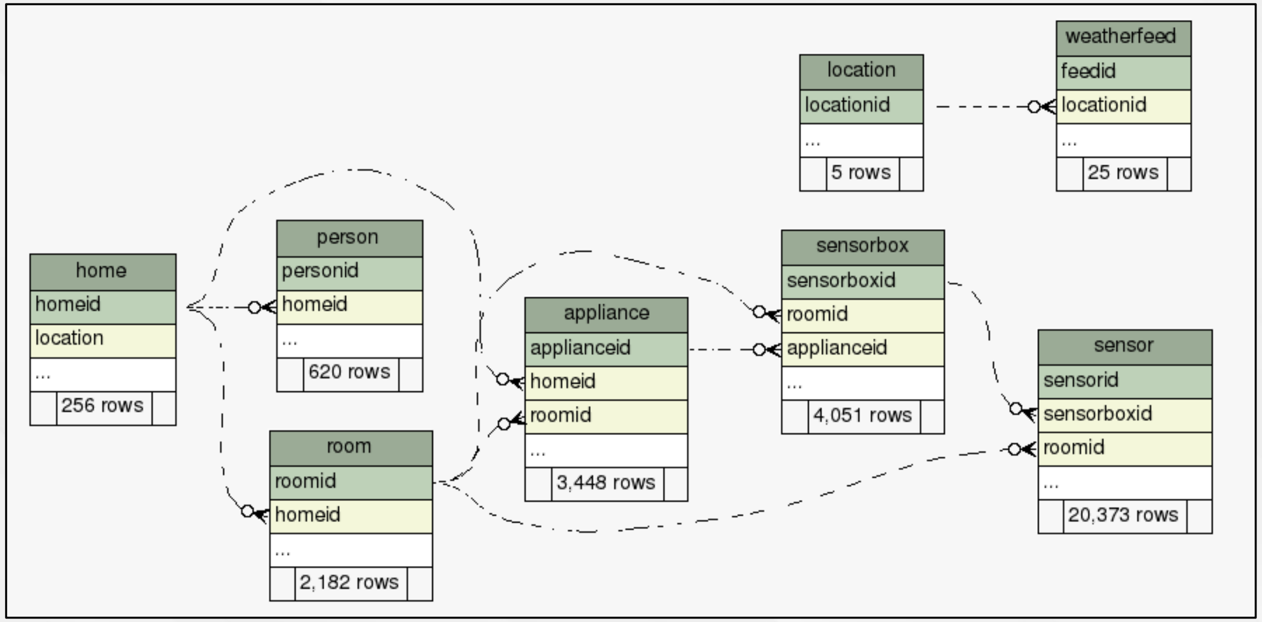
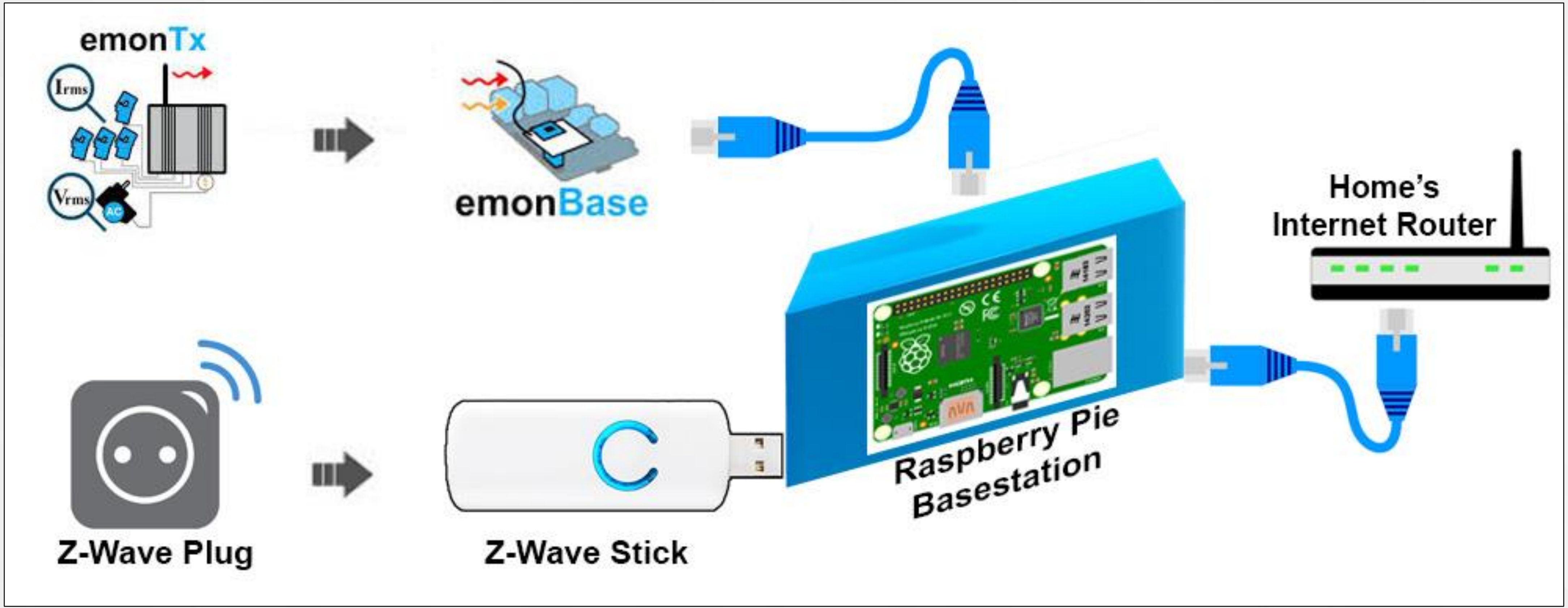


Diagram of relationships among the meta-data tables

Number of data elements belonging to 39 houses

Data Elements		Number
Rooms		354
Total number of occupants		87
Sensor boxes		1190
Sensors	Total number of sensors	5470
	Radiator sensors (sensor pairs per radiator)	2*209
	Lighting sensors	802
	Electric sensors for appliances	222

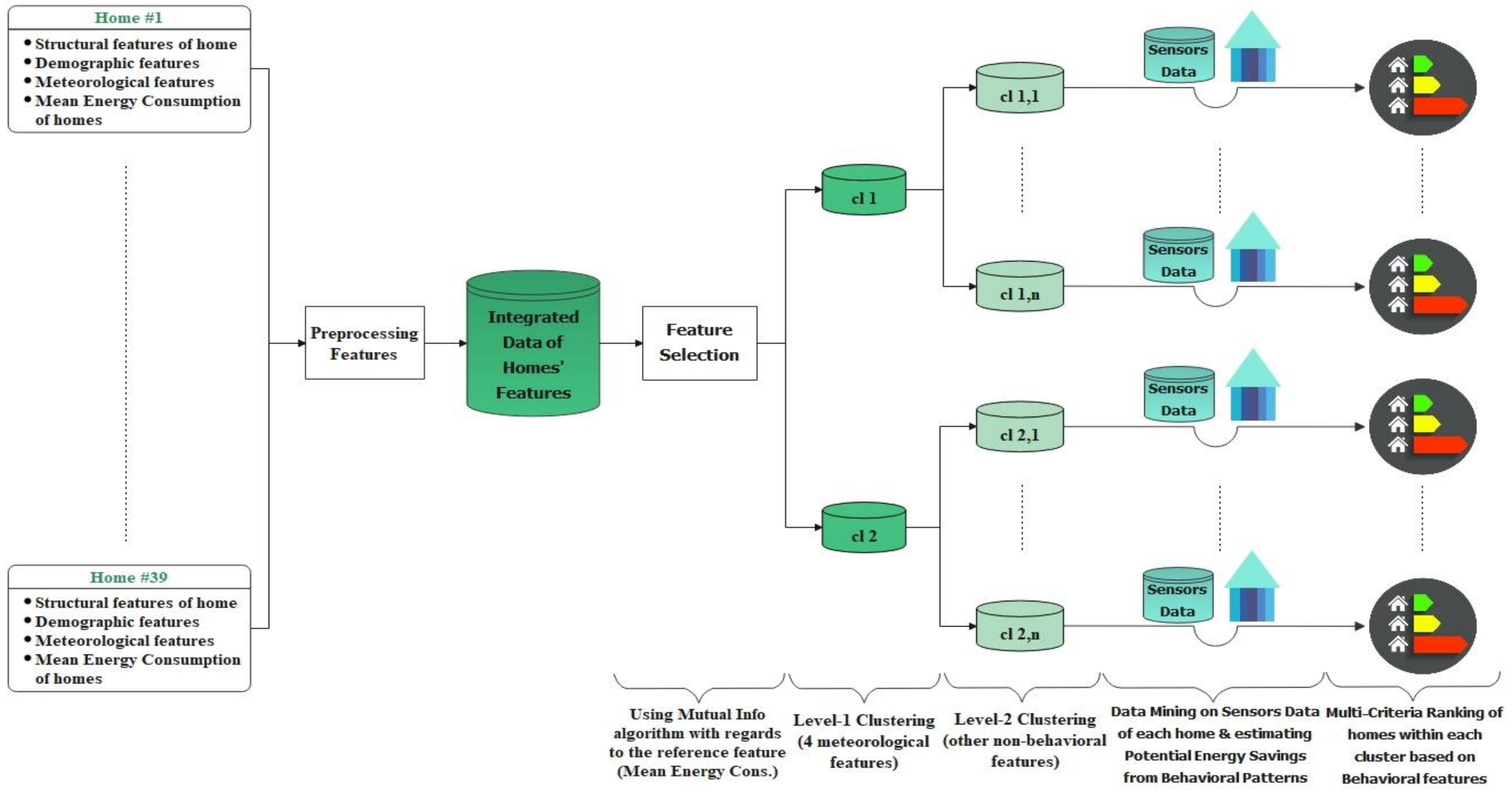
Data Collection Method



A schematic view of the combined Wireless Sensor Network used for the dataset

Model Workflow Overview

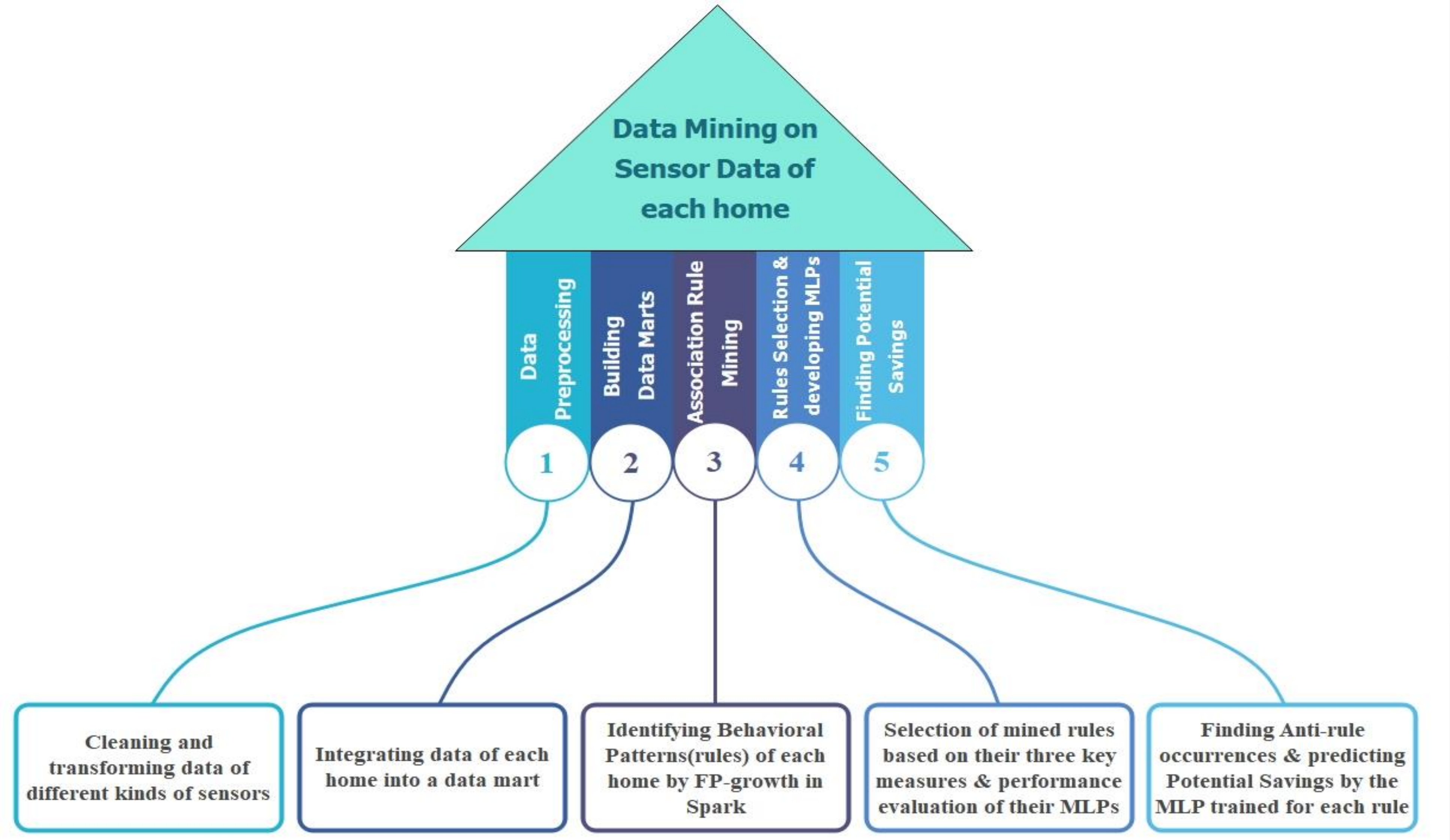
Method



A flowchart of the Model Stages and Procedures

Model Workflow Overview

Method



A flowchart of the data mining on the sensor data of the homes

Representing consumption behavioral patterns by association rules

Method

To briefly explain the process, the numeric consumption data for all devices across all timestamps is fed into our FP-growth model. This data is used to generate association rules, which reveal frequent energy consumption patterns of the appliances, reflecting occupants' behavior.

In the following paragraphs, we assume A, B, and C represent three appliances, such as electrical devices or radiators. The given hypothetical symbols represents two frequent energy consumption patterns for these appliances, as rules derived from the FP-growth model. Depending on the on/off state of the consequent (right-hand device status), the FP-growth model may yield two types of behavioral rules:

1. **Corrective Rule:** ***[A(on/off), B(on/off), ...] ---> [C(off)]***

   In the extracted rules like the one above, where the consequent device is mostly turned off, the pattern is identified as a "corrective rule." This implies that, if the consequent status is violated in subsequent events, a warning will be issued to the occupants to correct their exclusive energy-wasteful behavior.

2. **Reinforcing Rule:** ***[A(on/off), B(on/off), ...] ---> [C(on)]***

   In contrast, as shown above, if the consequent device is mostly turned on in the dataset, the pattern is classified as a "reinforcing rule." This implies that, if the consequent status is violated in subsequent events, an encouraging message will be sent to the occupants to reinforce their exclusive energy-conservative behavior.

The results of feature selection

The results of feature selection using the **Mutual Information** method, in order of the highest weights

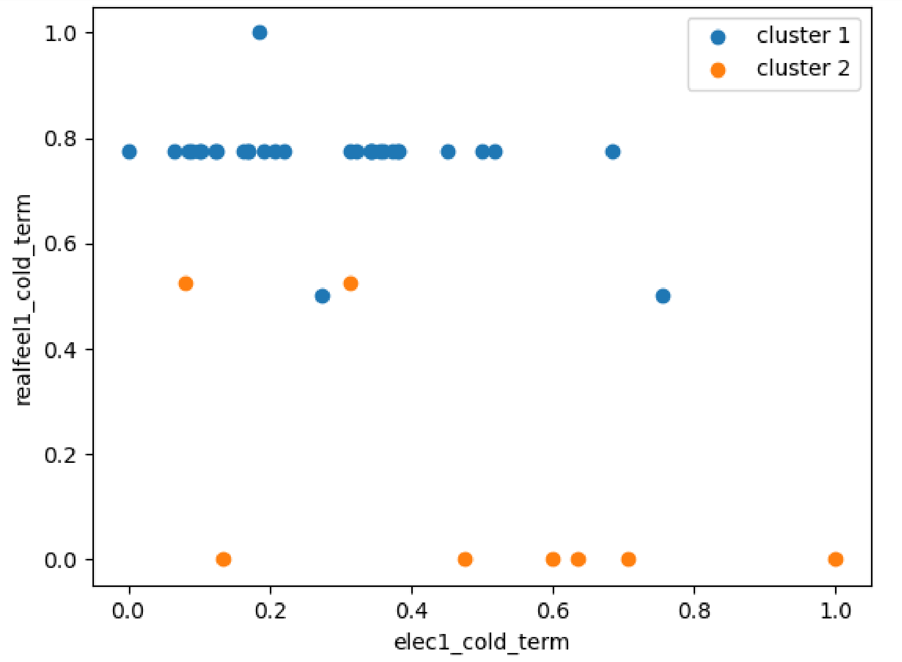
Buildings' Features	**Calculated weights**
Area of homes	0.161342
Length of residence per week	0.137747
Mean Relative Humidity (Meteorological)	0.113428
Number of households	0.113365
Feel temperature (Meteorological)	0.106375
Mean temperature (Meteorological)	0.101736
….	….

(Please note: Due to publishing restrictions on the results, only a subset of the top features is presented above.)

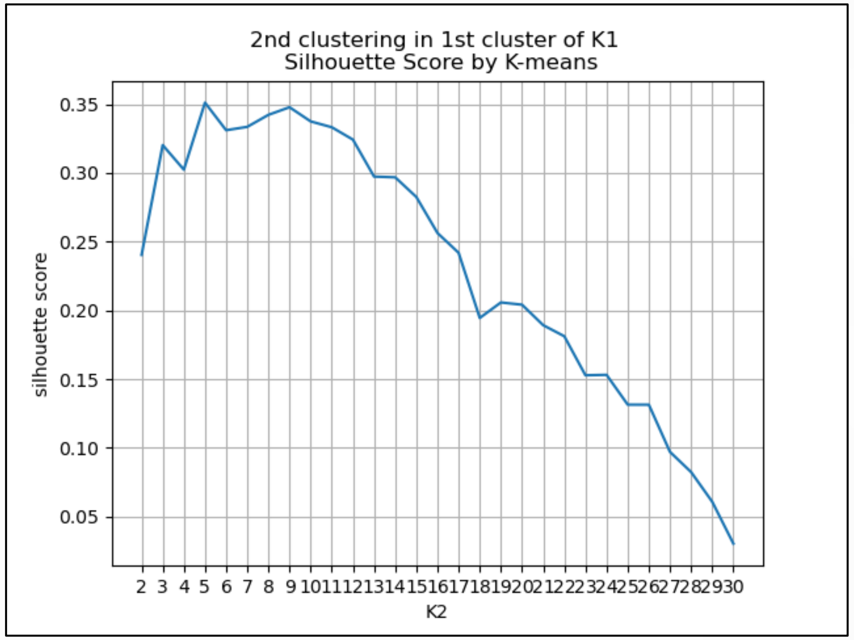
Two-level Clustering

Results

First some methods like One Hot, were used to convert qualitative features into quantitative forms. In the first stage of clustering, homes are clustered based on their four meteorological features, which have the most significant impact on their structural patterns for energy consumption. Based on the prominence of two clusters for tropical and cold climates, homes were divided accordingly. However, for the second level clustering we used Silhouette scores analysis for determination of cluster count in the resulting clusters of first level.



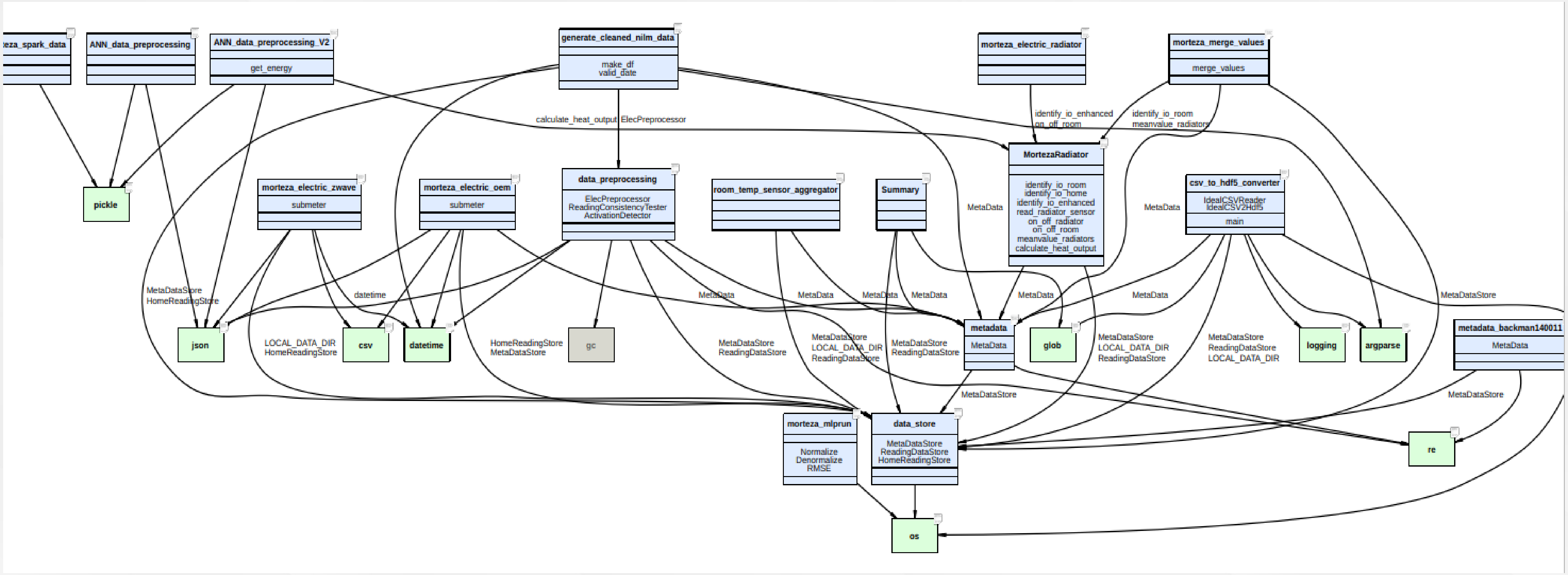
Scatter plot of homes based on Temperature and Average Energy Consumption (reference feature in MI method) according to the first stage of clustering.(tropical/cold climates)



Silhouette scores for various second-stage cluster counts among members of the first cluster from the initial clustering stage, aiding in the determination of the optimal number of clusters.

Results

A perspective of the complexity of mining sensor data



A graph of the relationships between Classes\ Methods\ Functions among more than 14 modules regarding only to the mining of the sensor data.

The Results of the innovative power on/off detection of appliances

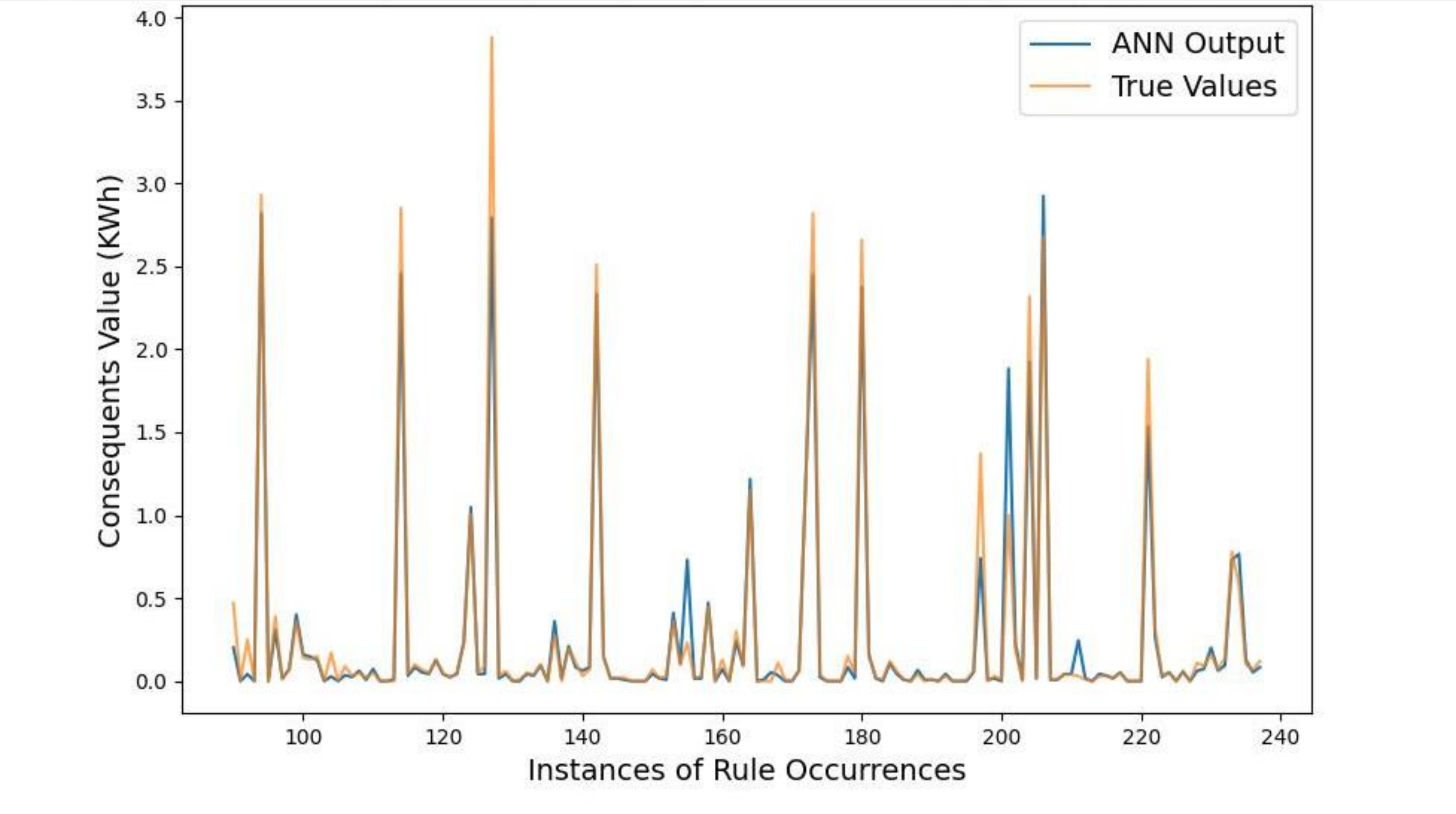
Results

An example of the performance of the power on/off detection module examining on sensor data of the tester appliance for home 65 using the innovative method of this study.

Row Num.	Time Stamp	Value	Time Duration (Calculated)	Operational Status (Calculated)
1419	6/9/2018 7:04:08	0	1144	0
1420	6/9/2018 7:23:12	1379	1	1
1421	6/9/2018 7:23:13	1379	1	1
1422	6/9/2018 7:23:14	1898	134	1
1423	6/9/2018 7:25:28	936	22	1
1424	6/9/2018 7:25:50	74	1	1
1425	6/9/2018 7:25:51	74	1	1
1426	6/9/2018 7:25:52	0	200	0
1427	6/9/2018 7:29:12	1883	122	1
1428	6/9/2018 7:31:14	944	22	1
1429	6/9/2018 7:31:36	248	2	1
1430	6/9/2018 7:31:38	0	1948	0
1431	6/9/2018 8:04:06	0	2412	0

(Please note: Electricity consumption of devices like Toaster was recorded by Z-Wave-based WSN, so the difficulty of handling these cases is about their ***variable sample rates***.)

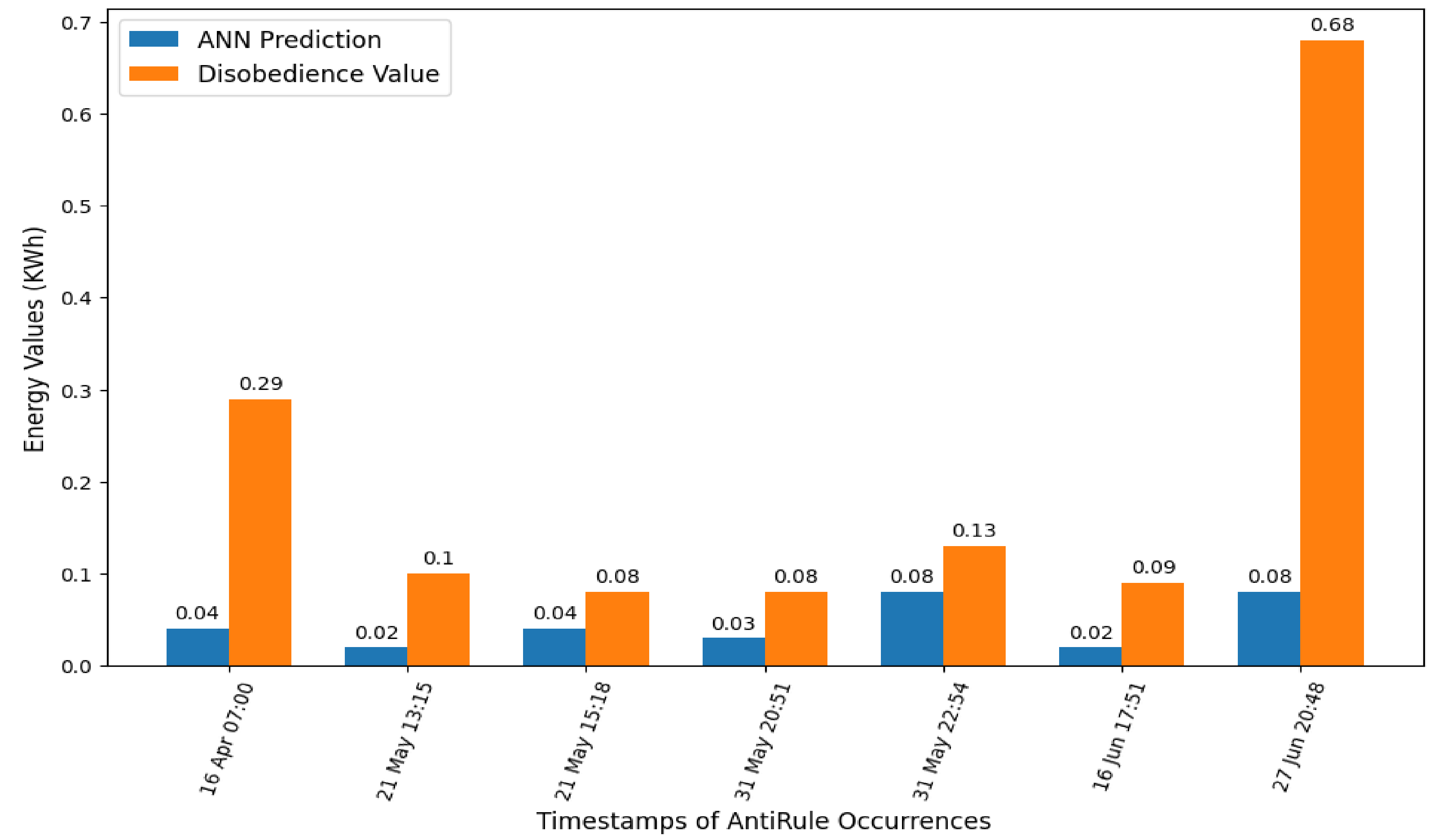
Training Multi-Layer Perceptrons (MLPs)



A presentation of the outcome from training an ANN for the rule [10, 350, 70] → [330] in home 145

Using MLPs for calculating Energy Saving Potential values

Results



The differences between the neural network predictions and the actual values in anti-rule events for the rule [10, 350, 70] → [330] in home 145 which imply the potential saving

Results

Calculating the energy saving potential of each rule by the MLPs

Output of ANN computations for some rules of House #106

Encoded form of the rule	Cross-validation error	RMSE	Selected Learning Rate from the Grid Search	Selected number of neurons in the hidden layer from the Grid Search	Type of the rule (Corrective vs. Reinforcing)	Type of consuming energy (in the Consequent)	Energy saving potentials from anti-rule occurrences
[10, 30, 100, 320]	0.115738	0.14463	0.001	50	Corrective	Gas	2.817928
[10, 320, 100, 330]	0.069352	0.06272	0.01	50	Corrective	Gas	1.808793
[10, 320, 70, 100]	0.096392	0.13715	0.001	40	Corrective	Electricity	0.063821
[10, 320, 70, 330]	0.151331	0.15443	0.01	40	Corrective	Gas	0.720388
[10, 320, 100, 340]	0.059531	0.05289	0.01	50	Corrective	Gas	1.445585

Ranking of homes based on the behavioral features within each cluster

The considered behavioral features for ranking homes within each cluster are as follows:

1. Energy saving potential
2. Mean Indoor-outdoor temperature difference
3. Electricity consumption per capita
4. Gas consumption per capita

Results

Ranking the building based on consumption behavior criteria in each cluster

After calculating the energy saving potential in each building in the last quarter of the dataset period (April to June, 2018), the entropy and weight of each criterion were calculated by using Shannon's method to rank the buildings in each cluster. Then, in TOPSIS, after calculating positive and negative ideals of each criterion, the similarity index for each option was obtained. For instance, in the following diagram, the result of ranking using this method is presented for the buildings in cluster K1=2, K2=2, where House # 259 with the highest similarity index ranks first (greatest consumption behavior performance), while House # 146 with the lowest similarity index ranks fifth (poorest consumption behavior performance).

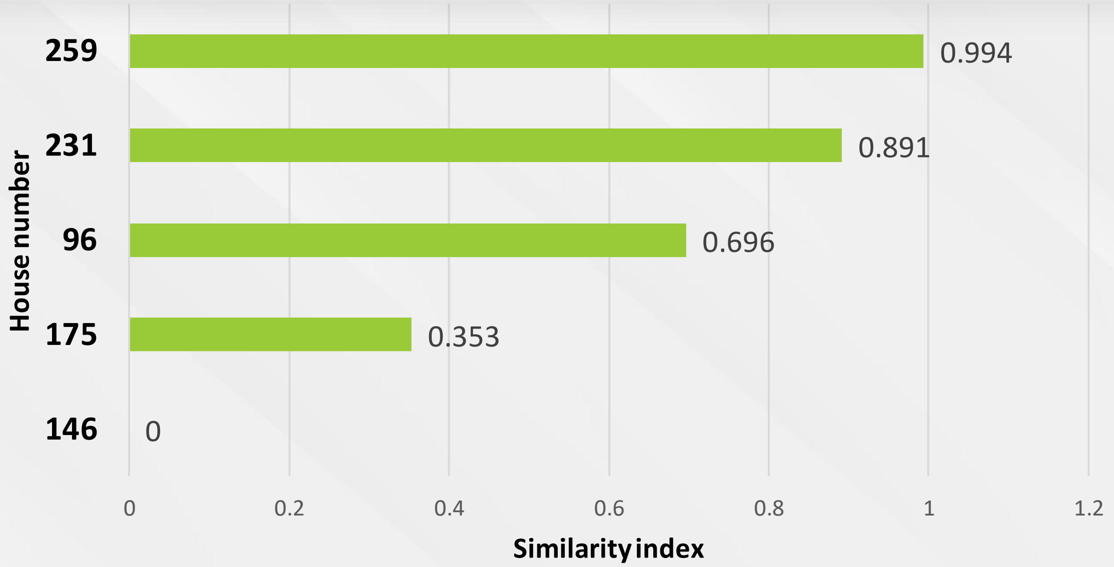


Diagram of the final ranking of options in cluster K1=1, K2=1

Ranking Results

Results

Ranking of houses in the cluster K1=1, K2=2 using the salient features of the cluster members

House number	TOPSIS score in the cluster	TOPSIS rank in the cluster	Gas consumption per capita (behavioral)	Electricity consumption per capita (behavioral)	Energy saving potential (behavioral)	Indoor-outdoor temperature difference (behavioral)	Number of occupants (non-behavioral)	Area of the house (non-behavioral)	Number of households (non-behavioral)	Year of construction (non-behavioral)	Length of residence per week (non-behavioral)
212	0.896	1	519.66	348.04	357	11.163	3	87.5	1	1909	12
139	0.726	2	2596.92	169.47	487	11.838	3	82	1	1874.5	12
255	0.66	3	2564.5	464.84	572	11.311	3	112	1	1874.5	14
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
162	0.03	8	5596.79	1356.47	466	13.859	2	73.5	1	2016	10
Average of features in this cluster:											
----	0.515	4.5	3179.84	689.2	426	12.27	2.5	83.25	1	1886	12.25

Discussion of Ranking Results

Results

Based on the rankings of houses in cluster K1=1, K2=2, the following conclusions can be drawn:

1. Ranking based on the four behavioral criteria of TOPSIS did not necessarily lead to an identical trend between the ranks and one of the four criteria; of course, this principle is a sign of the implementation of multi-criteria ranking, contrary to single-criterion ranking.

2. The similarity of all the items in terms of the five non-behavioral selected criteria, which had high weight in the Mutual Information method, is another evidence for the homogeneity of the structural properties of the houses in the final clusters – which is the objective of two-stage clustering. For instance, about 30% of the 39 houses had a single occupant; in this cluster, there is neither a house with one occupant, nor a house with more than 3 occupants. There is also a striking similarity in terms of the year of construction of the houses.

3. However, as for the behavioral criterion of indoor-outdoor temperature difference, only houses with 10 and 11-degree temperature differences were among the top four ranks of this clusters. In fact, choosing a warmer comfort temperature, which increases this criterion, reduces the rank of the houses in terms of consumption behavior in their clusters.

Discussion of Ranking Results

4. Another interest point relates to House #249 which ranked fifth. Although this house had the lowest energy saving potential (which is an advantage due to the negative sign of this criterion in this ranking), it also has the greatest indoor-outdoor temperature difference (about 14 degrees). In fact, one may expect that a house with the lowest energy saving potential (which shows optimal behavior and the lack of numerous anti-rules) must rank high in its cluster; however, due to choosing a warm indoor comfort temperature and, thus, high per capital gas consumption (about 900 kWh more than the cluster average), the consumption behavior of House # 249 cannot be evaluate as optimal from other aspects. This comprehensive view of the consumption behavior of houses was enabled by the Shannon-TOPSIS method.

Limitations

1. A limitation of this study was not providing precise data of the heating system. (e.g., the model or combustion efficiency of the combi boilers)
2. The data description related to the dataset of the study mentions that certain houses possessed high-consumption appliances such as A/C, computers, fans, etc. However, due to technical reasons or the occupants' unwillingness, no information was provided on the consumption of these appliances at this step.
3. Connecting to the automation system of houses that have such a system would provide useful information which could have improved behavior analysis and consumption recommendations, e.g., providing information on the temperature set by the occupants and awareness of window Open/Close sensors, which are useful in research in this domain.

Limitations

4. One potential area for improvement in this study could be expanding the sample size of houses to increase the reliability of certain inferences, particularly those discussed under "Discussion of Ranking Results". For instance, the small variance in the age of buildings is a consequence of the limited number of homes.
5. Although the large volume of sensory dataset in this research presented some Big Data challenges such as coping with HDF5 and Spark, the selection of 40 GB of data using a precise data pipeline still provided unique opportunities for analysis. For instance, analyzing the selected data with a small sample rate and high data granularity led to more sensible findings based on the literature. However, in comparison to controversial energy dataset in the literature works, the large volume of sensory dataset may have reduced the researcher's ability to process more methods and ideas on the dataset.

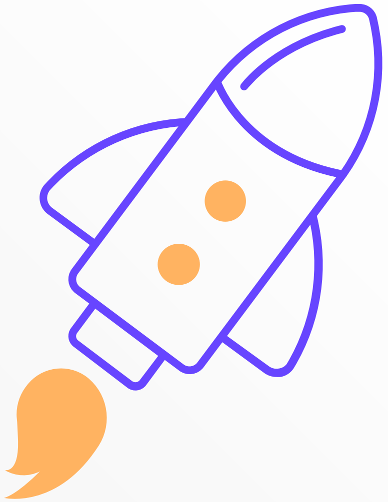
Suggestions

Kharazmi University

1. Utilizing data engineering tools such as AWS Data Wrangler, which offers unique interaction with the Pandas library and can accelerate processing, could be helpful in overcoming big data challenges in this study.
2. While the Discussion section introduces and discusses the behavioral phenomenon of Rebound Effect, future research could further explore potential solutions to this issue and examine to what extent the competitive ranking of houses might affect this misbehavior.
3. Despite being conducted in line with demand-side management policies, this thesis does not directly discuss a major method in this domain: Peak Shaving. However, new opportunities for this study could arise by solely having access to the “real-time price” of energy delivered to houses or the grid's variable energy tariff, to utilize Peak Shaving solution.
4. Building on the third suggestion, a promising research area could involve exploring the use of generated or stored electricity arbitrage with the local electricity grid, especially if more information on EV chargers or small-scale generators is provided in the dataset.

Appendix A:

Design and fabrication of a practical example of data collection tool by implementing a Wireless Sensor Network for home appliances energy usage



Brief explanation on the fabricated Wireless Sensor Network

Appendix

This appendix instructs the fabrication of a small-scale wireless sensor network that can prepare a dataset similar to the one used in this thesis. Note that the need for full access to more than 30 houses with the equipment required for each sensor network increased the time and costs of collecting such a large amount of data beyond a point that would be possible for a student without external support. Based on the esteemed examiner's comment, to complement this research, this sensor network is proposed through comparison with similar cases. Moreover, a practical example was constructed by connecting to a base station (based on Raspberry Pi 4B), fabricated with temperature and current sensor node (similar to the data of this study), and tested.

Brief explanation on the fabricated Wireless Sensor Network

Appendix

An image of a fabricated prototype of sensor nodes with ESP-12E WiFi Module, DHT11 sensor, and ACS712 sensor

When it comes to choosing the communication protocol, although it is still possible to use wired protocols such as KNX due to their lower costs, in the near future, few people may accept this cost and the risk of re-wiring the houses. Among the available options, a suitable choice for this project is inexpensive ESP8266 chips that allow flashing to professional users; however, their simple installation and operation (as will be explained) is still possible. Moreover, due to the consumption level of the WiFi protocol, they also have decent ability in utilizing the battery module.