

## Analysis of the Efficiency-Energy with Regression and Classification in Household Using K-NN

Qi Liu<sup>1,2</sup>, Zhiyun Yang<sup>1</sup>, Xiaodong Liu<sup>3</sup> and Scholas Mbonihankuye<sup>4,\*</sup>

**Abstract:** This paper aims to study energy consumption in a house. Home energy management system (HEMS) has become very important, because energy consumption of a residential sector accounts for a significant amount of total energy consumption. However, a conventional HEMS has some architectural limitations among dimensional variables reusability and interoperability. Furthermore, the cost of implementation in HEMS is very expensive, which leads to the disturbance of the spread of a HEMS. Therefore, this study proposes an Internet of Things (IoT) based HEMS with lightweight photovoltaic (PV) system over dynamic home area networks (DHANs), which enables the construction of a HEMS to be scalable reusable and interoperable. The study suggests a technique for decreasing cost of energy that HEMS is using and various perspectives in system. The method that proposed is K-NN (K-Nearest Neighbor) which helps us to analyze the classification and regression datasets. This paper has the result from the data relevant in October 2018 from some buildings of Nanjing University of Information Science and Technology.

**Keywords:** Energy management system, integrated wireless technology, home energy storage, home automation.

### 1 Introduction

In our life, electric energy plays a very important role to make efficient most of our activities. We always need electric energy because with evolution of the technology, most of the needs in daily life require energy and then the problem of energy management becomes one of the crucial problems that the humanity is facing to especially reducing electricity consumption and its cost. Heating, insulation, consumption, savings are some of the basics to consider while applying electrical devices in one's house. It is important to learn how to reduce your cost bill in order to gain your comfort life with less electric consumption and low electric energy cost. All circumstances or time which we need for accredit energy, we have to use accession of internet on the things devices and rising

---

<sup>1</sup> School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China.

<sup>2</sup> Shandong Beiming Medical Technology Ltd., 777 Shunfeng Rd, Jinan, 250101, China.

<sup>3</sup> School of Computing, Edinburgh Napier University, 10 Colinton Road, Edinburgh, EH10 5DT, UK.

<sup>4</sup> Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science & Technology, Nanjing, 210044, China.

\* Corresponding Author: Scholas Mbonihankuye. Email: Scholas.mbonihankuye@yahoo.fr.

numbering of contrivance acute in the house. The internet of things in energy usage is very significance in the devices convenience on our house [Kamoto, Liu and Liu (2017)]. The Home electric energy system has referred to some hardware and/or software, algorithm that can provide feedback about our necessity at home, and/or also enable advanced control energy-using devices at house (Working Group). Durability issues represent the biggest challenges that society is confronted. Least 83% to the world's most, electric energy, renewable energy (untenable) fossil meteorological, solar energy, and biomass is only about 2% sum above [Kolter and Johnson (2011)]. In KNN, the distance of each test data point to all neighbors is calculated and ranked in rising command. The top 'k' distances are taken, and the most frequent class in that subset used to define the class of that data point. When two element compound and  $\alpha$  is picket out correctly, the last score is appointed to minimize a total distance between those two elements, whereas the party of series had done in minimal number possible of bunch [Kamoto, Liu and Liu (2017)]. This is repeated for all data points until all have been labelled.

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} \quad (1)$$

$$\text{Standardized distance: } X_z = \frac{X - \text{Min}}{\text{Max} - \text{Min}} \quad (2)$$

### **1.1 Problem**

Energy consumption in house is significantly increasing in world. The first thing to understand is to know why energy consumption increasing quickly? Secondly is to determine what or when the troubleshooting electrical problem arise in a house. After, identifying the cause of troubleshooting problem system, we may choose the way that we can use for decreasing energy electricity.

### **1.2 Objective**

This study gives a systematic review with existing software, hardware, and confluence systems for managing electric energy in houses. Some of the tools which are often used are accessible on the bazaar and commonly used in daily life. The study gives an overview on methods used related to the position of smart home technologies by arguing main pertinent features, pointing out the strength of everyone method, the weakness of each technology and products used in order to regulate energy efficiency. The study will serve as the handbook for the users who want to opt technology that suits on their necessary, especially in reducing household electric equipment consumption.

### **1.3 Scope**

We aim to make a deep analysis of the problem related to home energy management system, and get better understanding of electric energy management based on result of our analysis, relying on efficiency-energy with Regression and Classification in household.

## **2 Related work**

Different methods or strategies used in electric energy sector in order to decrease energy consumption of household equipment are investigated. Electricity consumption is steadily

increasing since the 1990s, lately emerging as the second most used source of energy with a share of 17.7%, only behind oil with 40.8%. One of the leading factors for this growth in electricity demand is the change in the habits of energy consumption in domestic environments. In 2010, domestic consumption was responsible for 28% of the total electricity consumption among all sectors with an effective increase of 40% between 1990 and 2010 [Pereira, Quintal, Goncalves et al. (2014)]. It should be accommodated to mention that the potency economies of cooperation in communications have not been exploited in previous works. In the literature of wire-less communication base stations (BSs) to sleep through the thin delivery of traffic was the common draw near to save energy [Jamasandekar and Mudholkar (2015)]. The techniques done by Medasani and Kim shows that eliciting membership function form data are one of the fundamental applications with Fuzzy set theory [Medasani and Kim]. In the research done by Jamsandekar and Mudholkar, they proposed an algorithm such as fuzzy inference system (FIS) for classification of electric energy data. In their proposed work, they used a Genetic algorithm (GA) which is an optimal searching technique used for generation rules [Jamasandekar and Mudholkar (2015)]. The reference energy disaggregation Data Set aeration (REDD) [Kolter and Johnson (2011)] presents a data set containing freely available power information on the use of several houses, which aims to advance search on energy ventilation. These kinds of data are used in testing or implement methods that can be applied to HMES in order to reduce energy consumption.

Soumadip Ghosh et al. have proposed a genetic algorithm approach for finding frequent item sets which caters to positive and negative association mining [Medasani and Kim (1998)]. In their result, they showed the rule of mining problem to find frequent item sets using their proposed GA based method, they found that it is a very simple and efficient one.

Marmelstein proposed the method in which he tried to explore the different methods of using genetic algorithm with K-Nearest Neighbor algorithm to improve the classification accuracy and minimize the training error [Ghosh, Biswas and Sarkar (2010)]. Soraj et al. proposed a Genetic fuzzy logic method for the discovery of decision rules of datasets containing categorical and continuous attributes [Marmelstein (1997)]. From the research done by Liu and Zhu, they found that among 2010 to 2030, the save energy in-creasing should be imprinted to a global energy require twice more than the one known today. According to their analysis and view the rate of the high technology today, their findings are about to be correct. Figure out these challenges, many modifications should be made regarding the existing electrical energy system [Parbhat (2011)]. Considering the findings from the work done by Liu et al. [Liu, Qiu, Fan et al. (2016)], the alteration of cli-mate had influence on energy electricity. Generally, communication networks and data centers are known to be the largest power consumers and Green House Gases (GHG) emitters among other ICTs, benefit from smart grid driven techniques to enhance energy savings and emission reductions [Liu, Qiu, Fan et al. (2016)] for a sustainable development.

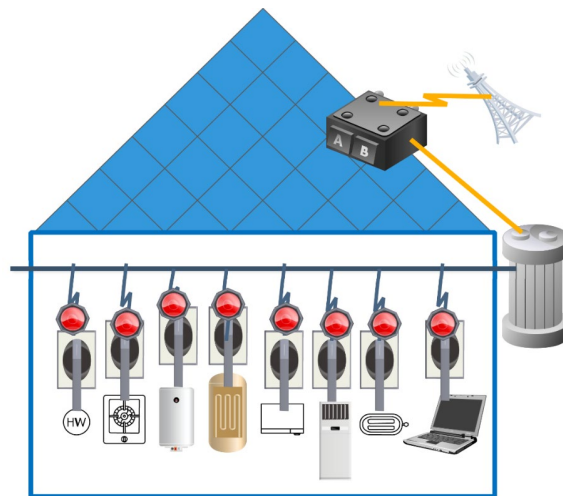
### **3 Data and methodology**

In order to efficiently conduct our study, we used data collected from Nanjing University of Information Science and Technology, International students' dormitory number 4, room 322. The data collected concerned energy consumption such as telephone, heat

water, heat water dispenser, air conditioner, iron. To achieve the goal on studying energy consumption in household, we used K-Nearest Neighbors (KNN) which is a non-parametric method mainly based on the analysis about big data. It is used for classification and regression to clarify the space among each point which is identified in clustering. The result obtained depends of the value gave by k, where K-NN algorithm used for classification or regression the distance. This method has been used in order to study how to reduce energy consumption in a household. In KNN, the distance of each test data point to all neighbors is calculated and ranked in ascending order. The top 'k' distances are taken, and the most frequent class in that subset are used to de-fine the class of that data point. That step is repeated for all data points until all have been labelled.

### ***3.1 Appliance modeling***

Regression models are one of the simpler yet powerful analysis methods for understanding relationships in the data and generating predictions from them. This is normally done using the Least Squares Method, which attempts to fit a 'line of best fit' that minimizes the sum of squares of the vertical difference of each point from the line itself [Erol-Kantarci and Mouftah (2015)]. K-Nearest Neighbors (KNN) is the most common algorithm used. It's a supervised learning technique, where given a data point, the algorithm will output a class membership for that point [Rogozhnikov (2015); Yao, Li and Kumar (2010)]. KNN can also be used for identifying outliers in data. The Fig. 1 shows a system which is com-posed by many devices used to get electric energy. Energy electricity system flows into the lights, appliance from the socket to our devices. We have tried a schematic Fig. 1 with some devices of home energy electricity system.

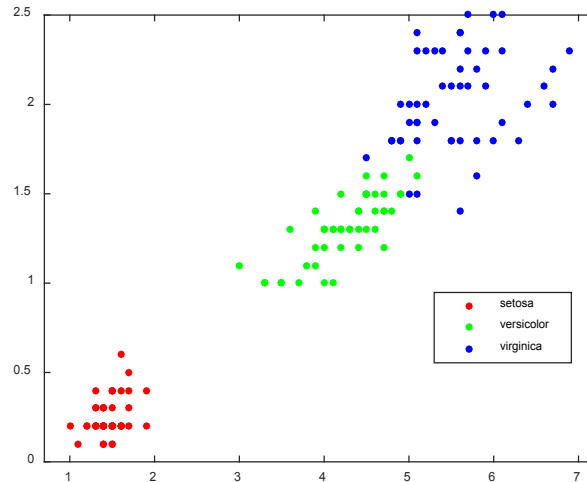


**Figure 1:** Formulary system house layout

### ***3.2 KNN classification and regression works***

This method is used for processing of classification and regression in a given dataset such as electric energy analysis. The Kernel-Nearest Neighbor (KNN) normally use the

classification and regression on their functionality. The output is a category of members. An object is classified by a majority votes by majority adjacent with the object assigned to the most popularly class in the K most related. Its neighbors can also be used for regression. Output value is obtained using production rerun.



**Figure 2:** Classification and regression

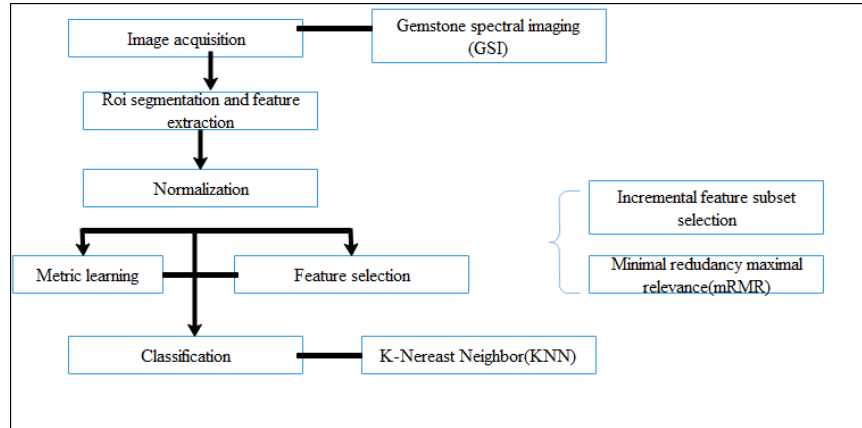
Steps by steps to computer KNN algorithm

- First of all, we determine the parameter k which is the number or proximity of nearest neighbor;
- Computer the distance between patterns;
- Spell outrun and firm the most proximity K-th based on firm minimal;
- Put together patterns by category based on proximity most than nigh;
- Majority sample of the category of nearby most nigh becomes values of prediction of the search. Weigh more similar house more than those less similar in list of K-NN.

**3.3 Algorithm overview**

We prefer to use K-Nearest Neighbor algorithm (k-NN) that is one of non-parametric method used for classifying data or regress data because it has been widely used by other authors and is also chosen to be one of the most important techniques in home energy management system analysis [Kamesh and Sakthi (2014); Kornaropoulos and Tsakalides (2009); Chen, Li and Tang (2013)].

$$\hat{Y}_{q=} = \frac{C_{qNN1}Y_{qNN1} + C_{qNN2}Y_{qNN2} + C_{qNN3}Y_{qNN3} + \dots + C_{qNNK}Y_{qNNK}}{\sum_{j=1}^K C_{qNNj}} \tag{3}$$



**Figure 3:** Overview of K-NN

For the static modeling regression analysis, it has been set a series of statistical processes in order to estimate the relationship among devices. K-Nearest Neighbor algorithm is a way to classify target with attributes to its nearest neighbor in the Learning set. In K-NN method, the K-Nearest Neighbors are considered [Bhatia (2010); Garcia and Gómez (2006); Zhang and Zhou (2007)].

Below are formulas used to calculate a distance function of the variable:

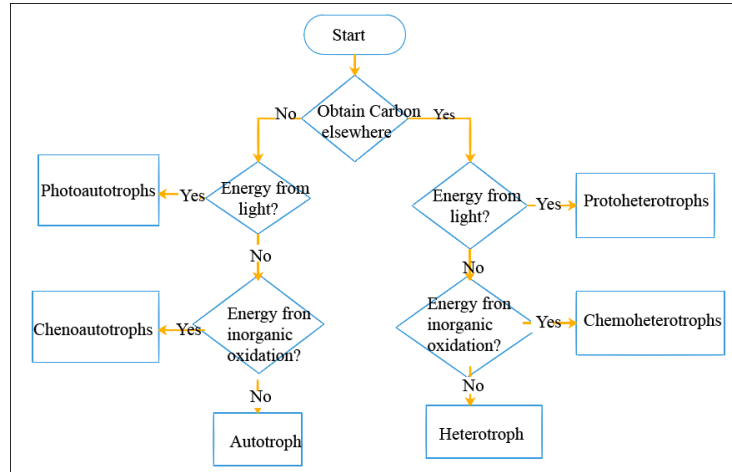
$$Euclidean = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (4)$$

$$Minkowski = \left( \sum_{i=1}^k (|x_i - y_i|^q)^{1/q} \right) \quad (5)$$

$$Minkowski = \left( \sum_{i=1}^k (|x_i - y_i|)^q \right)^{1/q} \quad (6)$$

$$Distance \ Hamming: D_H = \sum_{i=1}^k |x_i - y_i| \quad \begin{cases} x = y \Rightarrow D = 0 \\ x \neq y \Rightarrow D = 1 \end{cases} \quad (7)$$

The problem that we have been resolving in logistic regression, the variable need depends on binary variable (True/False, Yes/No), for example in classification problems. It can be highlighted that logistic regression can be used even if the dependent and independent variables in our model do not have a linear relationship.



**Figure 4:** Classification of problem regression

**3.4 Flow way (electricity flows)**

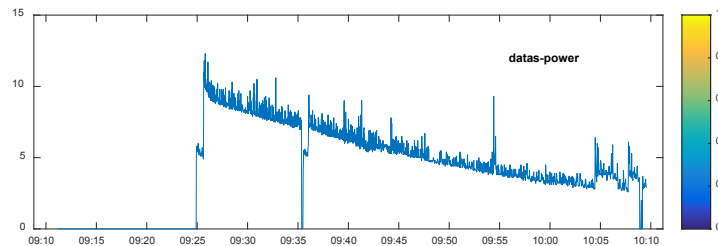
The way of electricity through the transformer is on the one hand and operation house items. If it touches a hot wire, provides the long way path for electricity to flow between wire and transformer terminal.

**4 Validation with experimental data from room 322, #4 building of NUIST**

The dataset used to validate our experimentation has been personally collected from someone’s room, East campus of Nanjing University of Information Science and Technology precisely old dorm room number 322 for a period of 3 days averaging from October 21th to 24th 2018. We tried to analyze different house-hold equipment and studied which devices use high energy consumption.

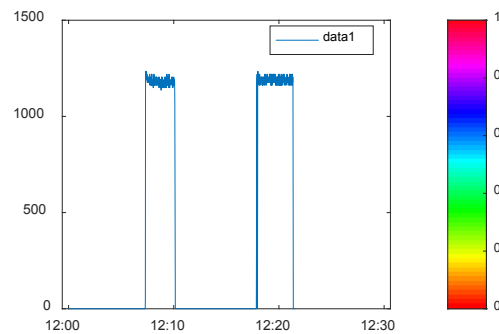
**4.1 Result found from experiment**

Most of our experimentations have been done using MATLAB software. Thus, all figures used in this study are obtained from the analysis of data electric energy consumption using MATLAB. The electricity was downloaded from different devices and from different sockets.



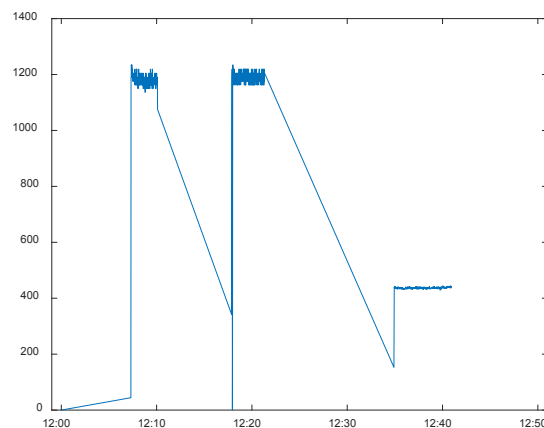
**Figure 5:** Curve of data \_ power

The Fig. 6 shows the variation of power consumption within each period of 10 min. Then we found that from 12:08 to 12:10 the power consumption is very high. The question that comes out is why the energy consumption was suddenly getting higher and higher. We noticed that, when someone is heating water or cook something to eat, the energy consumption becomes very high at the time when proceed to finish. During our experimentation after 12:10, it can be seen a sudden decrease like a latent period which resumed around 12:20. During that period, the power consumption became very high. The energy consumption follows a sinusoidal curve. The Fig. 6 shows the result obtained from the data collected on October 24th in 2018 in the room of the East campus in Nanjing University of Information Science and Technology.



**Figure 6:** Curve of data-power2

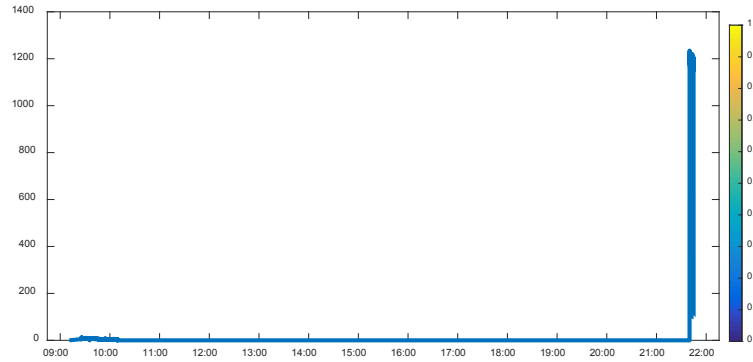
In the Fig. 7, we found that the variation starts after 10 min. The data were collected on October 24th, 2018. The result has shown that from the starting time 12:00, the variation starts after 10 min and we observe a sequence of increasing and decreasing phenomena of energy consumption. It can be seen that the heat water energy consumption differs from the previous device. However, at the end of the collecting period, the figure shows us that curve of variation still on the same level of temperature.



**Figure 7:** Curve of data-power3 (heat water)

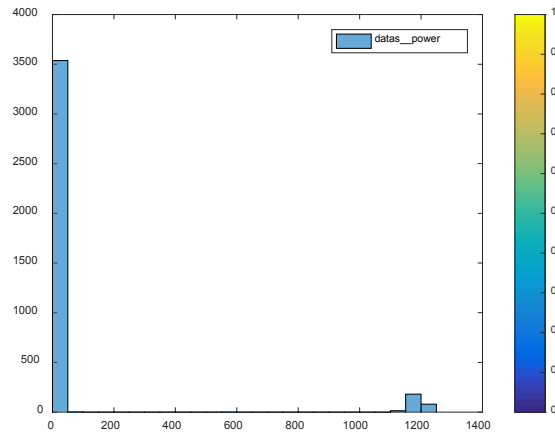


As shown in the Fig. 8, the electric energy consumption of power bank device is not very high. As shown in the figure most of the charging time the energy consumption is mostly closer to zero. However, at the end of the collection, considering the charging period, the power consumption increases very quickly. This data was collected on 21st October 2018.



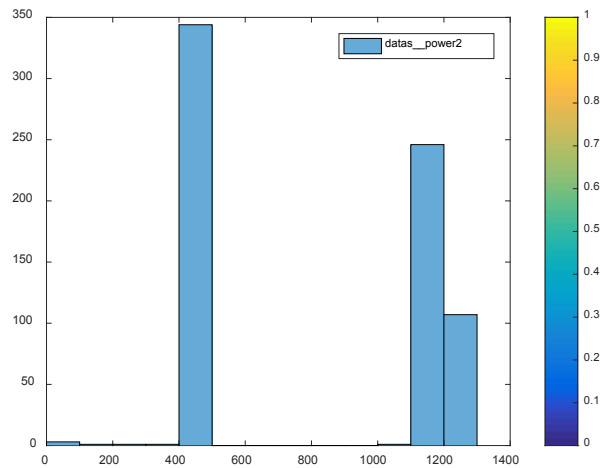
**Figure 8:** Curve of data-power4 (power bank)

The Fig. 9 presents a histogram showing the data of energy consumption of tele-phone when is charging. At the first time it registers a very high-power consumption, and suddenly, comes a period where the devices seem not to consume energy. However, the energy consumption resumes after a long period with a low consumption comparing to the starting period.



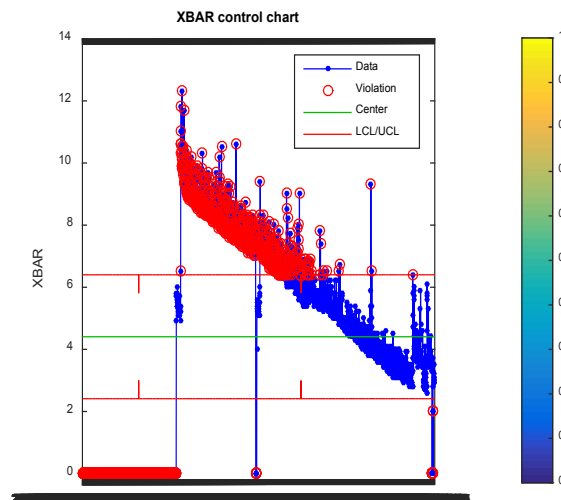
**Figure 9:** Histogram of charging telephone

The Fig. 10 shows a histogram obtained while measuring the heat water energy consumption. Three phases can be identified. At the beginning, the power consumption is low but around 40% of heating, the energy consumption becomes very high then decreases for a certain latent period. The energy consumption increases again at the end of the heat process when the water is at around 85% of its boiling level. At the end of the boiling phase there is no more power consumption.



**Figure 10:** Histogram of heat water

The Fig. 11 presents a XBAR control chart and show us how a process changes over time. Control chart many times has center with the result Fig. 11 we get violation data center and LCL/UCL. These lines are determined by historical store data used.

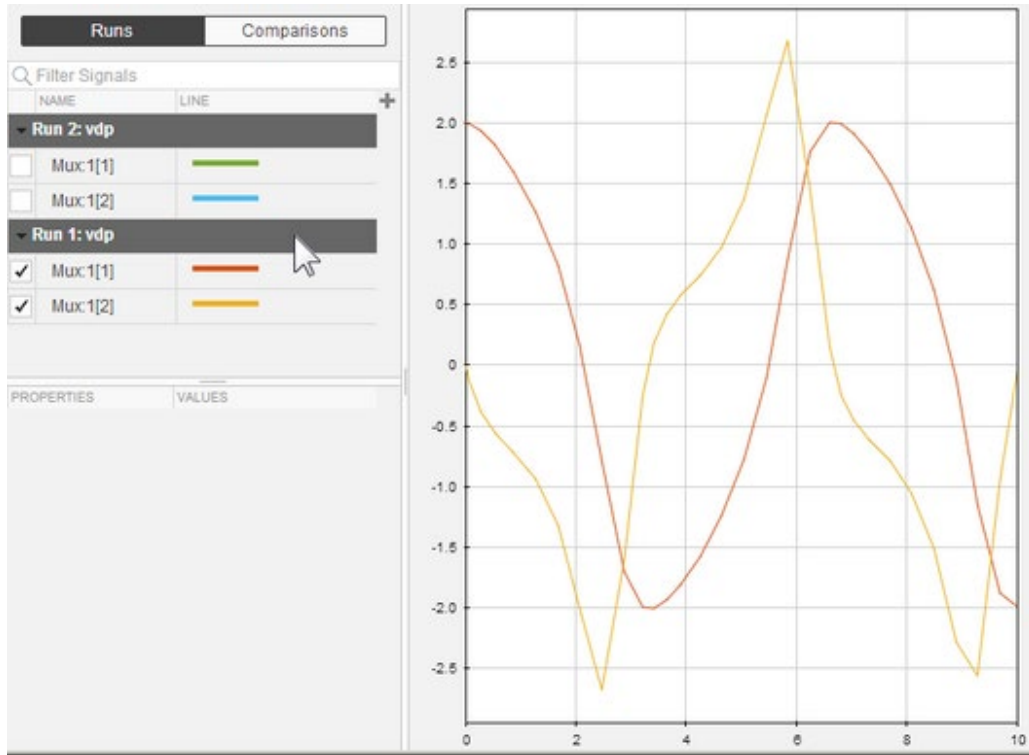


**Figure 11:** Control chart

#### ***4.2 Simulation and comparison***

In any home energy management system analysis, it is necessary to make a simulation of the data for a given house to be simulated to other houses in order to make a general and perspective analysis and establish some assumptions. This model simulation has been run with MATLAB software. The Data that we collected from the model during real-time execution or normal simulation, and stored to a variable gave us the schemes or figures of stream signal whom will be a result of our simulation. MATLAB workspace has some

functions very important when we need to do analyze of data or if we need the plotting functions in comparison about variation and for visualization of our purposes. The Fig. 12 shows the simulation output.



**Figure 12:** Stream signal from Simulink

## 5 Conclusion

Energy consumption and energy storage are among vital issues for human society and industry. For states, energy independence is strategic and economically essential. For individual and businesses, energy must be available on demand, without any sudden interruption. Any breakdown of energy supply has a high economic and social cost and negative impact in terms of health and safety. The Energy Storage Service will help to uncover the revenue streams and business opportunities most relevant to any project. It provides a foundation in the economics, market landscape and technology advancements that is essential to formulating innovative strategies in the energy storage market. It has been demonstrated that the energy consumption depends on the categories of electric devices, electrical appliance, the allocation and the number of electric equipment in the house. The energy consumption variation depends also on the way it is used. Some homes used more household electric equipment while others are using few electric equipment. We have shown that for example the telephone and heat water are not consuming the same energy.

**Acknowledgement:** This work has received funding from 5150 Spring Specialists (05492018012), College Students' Enterprise and Entrepreneurship Education Program of Jiangsu Province, the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 701697, Major Program of the National Social Science Fund of China (Grant No. 17ZDA092), 333 High-Level Talent Cultivation Project of Jiangsu Province (BRA2018332) and the PAPD fund.

## References

- Bhatia, N.** (2010): Survey of nearest neighbor techniques. *International Journal of Computer Science and Information Security*, vol. 8, no. 2, pp. 302-305.
- Chen, Q.; Li, D.; Tang, C.** (2013): KNN matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2175-2188.
- Erol-Kantarci, M.; Mouftah, H.** (2015): Energy-efficient information and communication infrastructures in the smart grid: a survey on interactions and open issues. *IEEE Communications Surveys and Tutorials*, vol. 17, no. 1, pp. 179-197.
- García, C.; Gómez, I.** (2006): *Algoritmos De Aprendizaje: Knn & Kmeans (Ph.D. Thesis)*. Carlos III University of Madrid, Spain.
- Ghosh, S.; Biswas, S.; Sarkar, D.; Sarkar, P. P.** (2010): Mining frequent itemsets using genetic algorithm. *International Journal of Artificial Intelligence & Applications*, vol. 1, no. 4, pp. 133-143.
- Jamsandekar, S. S.; Mudholkar, R. R.** (2015): Fuzzy inference rule generation using genetic algorithm variant. *Journal of Computer Engineering*, vol. 17, no. 4, pp. 9-16.
- Kamesh; Sakthi, P. N.** (2014): Security enhancement of authenticated RFID generation. *International Journal of Applied Engineering Research*, vol. 9, pp. 5968-5974.
- Kamoto, K. M.; Liu, Q.; Liu, X.** (2017): Unsupervised energy disaggregation of home appliances. *Lecture Notes in Computer Science. International Conference on Cloud Computing and Security*, vol. 10602, pp. 398-409.
- Kolter, J. Z.; Johnson, M. J.** (2011): A public data set for energy disaggregation research. *International Workshop on Data Mining Applications in Sustainability*.
- Kornaropoulos, E. M.; Tsakalides, P.** (2009): A novel KNN classifier for acoustic vehicle classification based on Alpha-Stable statistical modeling. *IEEE Workshop on Statistical Signal Processing Proceedings*, pp. 1-4.
- Liu, Y.; Qiu, B.; Fan, X.; Zhu, H.; Han, B.** (2016): Review of smart home energy management systems. *Energy Procedia*.
- Marmelstein, E.** (1997): Application of genetic algorithms to data mining. *Application of Genetic Algorithms to Data Mining*, pp. 53-57.
- Medasani, S.; Kim, J.** (1998): An overview of membership function generation techniques for pattern recognition. *International Journal of Approximate Reasoning*, vol. 19, no. 3-4, pp. 391-417.
- Parry, R. M.** (2010): K-nearest neighbor models for microarray gene expression analysis and clinical outcome prediction. *Pharmacogenomics Journal*, vol. 10, pp. 292-309.

**Pereira, L.; Quintal, F.; Goncalves, R.; Nunes, N. J. S.** (2014): SustData: a public dataset for ICT4S electric energy research. *Proceedings of the 2014 Conference ICT for Sustainability*.

**Prabhat, N.** (2011): A genetic-fuzzy algorithm to discover fuzzy classification rules for mixed attributes datasets. *International Journal of Computer Applications*, vol. 34, no. 5, pp. 15-22.

**Rogozhnikov, A.** (2015): Machine learning. *High Energy Physics*.

**Yao, B.; Li, F.; Kumar, P.** (2010): K-nearest neighbor queries and KNN-joins in large relational databases (almost) for free. *International Conference on Data Engineering*, pp. 4-15.

**Zhang, M. L.; Zhou, Z. H.** (2007): ML-KNN: a lazy learning approach to multi-label learning. *Pattern Recognition*, vol. 40, no. 7, pp. 2038-2048.