aurum

AURUM MÜHENDİSLİK SİSTEMLERİ VE MİMARLIK DERGİSİ
AURUM JOURNAL OF ENGINEERING SYSTEMS AND ARCHITECTURE

Cilt 6, Sayı 1 | Yaz 2022
Volume 6, No 1 | Summer 2022, 45-65

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

INTERNET OF THINGS BASED ZIGBEE SNIFFER FOR SMART AND SECURE HOME

Farah Shakir Mahmood ALBAYATİ[1]
[1]Electrical and Computer Engineering, Altinbas University, Istanbul, Turkey
farah.shakir91@gmail.com, ORCID No: 0000-0003-4257-4603

Galip CANSEVER[2]
[2]Electrical and Computer Engineering, Altinbas University, Istanbul, Turkey
galip.cansever@altinbas.edu.tr, ORCID No: 0000-0003-2294-4259

GELİŞ TARİHİ/RECEIVED DATE:19.08.2021 KABUL TARİHİ/ACCEPTED DATE:16.05.2022

Abstract

This paper aims to resolve the Internet of Things (IoT) based ZigBee sniffer for smart home and determine the usage of energy or power with high spectrum allocation in future ZigBee Protocol with the help of clustering in IoT with data mining. The research work starts presenting an overview of the broadband network energy sector and the challenges that face it. It is observed a change in the energy policies promoting energy efficiency, encouraging an active role of the consumer, instructing them about the importance of consumer behavior, and protecting consumer rights. Electricity is gaining room as an energy source. Its share will keep constantly increasing in the following decades. ZigBee Protocol and smart meters' deployment will benefit both the utility and the consumer in the near future. New services and new businesses appear in this environment, focusing on the energy management field and tools. They require specialization in fields such as computer science, software development, and data science. This research has segmented the ZigBee Protocol according to the similarities of their electrical load profiles, using the proportion of energy usage per hour (%) as a common framework. This energy consumption segmentation aims to provide personalized recommendations to each group to reduce their energy consumption and the associated costs, fostering energy efficiency measures and improving consumer engagement. The desired segmentation is obtained by an iterative process, based on computational clusters calculation (using a Python programming language) and finalized by a post-clustering analysis applying visualization and statistical data mining technique to detect the energy consumption and reallocate them to a more appropriate group. The K-Means clustering technique was tested and compared, giving the best prediction of accuracy 98.46% for all energy load profiles with a high spectrum of 100GHz. The solution from the K-Means clustering is the one that better adapts to the segmentation sought, which is used as the base of the post-clustering stage to obtain the final energy consumption segmentation.

Keywords: Energy, Internet of Things (IoT), K-Means, ZigBee Protocol, Clustering, Consumption, Smart, Monitoring.

AKILLI VE GÜVENLİ EV İÇİN ŞEYLERİN İNTERNETİ TABANLI ZİGBEE SNIFFER

Özet

Bu araştırmanın amacı, IoT'de kümeleme yardımı ile ZigBee Protokolünde yüksek spektrum tahsisli enerji veya güç kullanımını belirlemektir. Bu araştırma, geniş bant şebeke enerji sektörüne ve karşı karşıya olduğu zorluklara genel bir bakış sunmaya başlar. Enerji verimliliğini teşvik eden, tüketicinin aktif rolünü teşvik eden, tüketici davranışlarının önemini anlatan ve tüketici haklarını koruyan enerji politikalarında bir değişiklik gözlemlenmektedir. Elektrik, enerji kaynağı olarak yer kazanmakta olup, önümüzdeki on yıllarda da payı sürekli artmaya devam edecektir. Bu enerji tüketimi segmentasyonunun arkasındaki amaç, enerji tüketimini ve ilgili maliyetlerini azaltmak, enerji verimliliği önlemlerini teşvik etmek ve tüketici katılımını geliştirmek için her gruba kişiselleştirilmiş öneriler sunabilmektir. İstenen segmentasyon, hesaplamalı kümeler hesaplamasına (python programlama dili kullanılarak) dayanan yinelemeli bir süreçle elde edilir ve enerji tüketimini tespit etmek ve bunları daha uygun bir gruba yeniden tahsis etmek için görselleştirme ve istatistiksel veri madenciliği tekniğini uygulayan bir kümeleme sonrası analizi ile sonuçlandırılır. K-Means kümeleme tekniği test edildi ve karşılaştırıldı, 100GHz yüksek spektrumlu tüm enerji yükü profilleri için en iyi doğruluk tahminini %98.46 verdi. K-Means kümelemesinden elde edilen çözüm, nihai enerji tüketimi segmentasyonunu elde etmek için kümeleme sonrası aşamanın temeli olarak kullanılan, aranan segmentasyona daha iyi uyum sağlayan çözümdür. Bu metodolojilerin çoğu, daha yüksek enerji tasarrufu potansiyeline sahip kullanıcıları belirlemeye odaklandıkları için 100 kWh cinsinden mutlak değerleri kullanır. Bu durumda, enerji tasarrufu tavsiyelerinin ZigBee protokolünün belirli özelliklerine göre kişiselleştirilmesine, uygun zamanda yeterli tavsiyeyi sunarak tüketici deneyiminin iyileştirilmesine, enerji verimliliğinin etkinliğini artıran gerçeklere, geleceğe yönelik tavsiyelerin hizmetine izin verir. ZigBee protokolü

Anahtar Kelimeler: Enerji, Nesnelerin İnterneti (IoT), K-Means, ZigBee Protokolü, Kümeleme, Tüketim, Akıllı, İzleme.

1. INTRODUCTION

The Internet of Things can be defined as the network of physical devices embedded with electronics, software, sensors, actuators, and connectivity, enabling them to exchange data with each other and the internet using their unique identifiers as mentioned in (Becirovic and Mrdovic, 2019). Such an infrastructure allows the development of applications and services that facilitate the lives of citizens and workers around the world and bring us closer to the convergence of the physical and digital worlds. Some successful examples include Nest and Ecobee for smart-heating, Phillips Hue for smart-lighting. Smart things as a provider for multiple smart-home solutions, researchers in (Wu, et al., 2015) worked for personal fitness trackers and personal assistants like Alexa from Amazon or Siri from Apple.

Figure 1. A smart ZigBee used with several popular firms.

Despite the massive growth of IoT applications, the usefulness of the data originating from these systems is still to be validated and proved, as mentioned in (Bouktif, et al., 2018). The volume of data generated from a single sensor device installed in a house or a wearable device can be overwhelming for the device itself. In most cases, it needs to be offloaded to a data-processing application from which users can access it and connect it to a more meaningful use (Ge et. al., 2017). The rate of such devices integrated into our everyday lives causes offloading information in the cloud. Therefore, it creates vast volumes of data, as mentioned in (Björnson et. al., 2015). Such data need to be communicated, processed, and stored in real-time while accessing it in its raw format tends to be useless when compared with the knowledge it can generate (e.g., information about a person's lifestyle and diet compared with their daily step count and trips as latitude-longitude coordinates) as mentioned in (Aslan et. al., 2017). As a result, the need for a holistic solution on managing the data generated is still an open subject for research in both academic and enterprise scopes. Similarly, multiple interfaces for communicating and representing such data formats have been used, with none getting a clear step ahead of the others in the path for a global standard, if and when this will be possible.

1.1. Problem Statement

In order to provide such a holistic solution, it is essential to be able to integrate the data generated by all of these smart homes in a common base. We need to be able to process them, view them and operate on them using a standard methodology and also be able to gain common knowledge with security monitoring. This is the most critical step towards unifying the information from a fragmented ecosystem that significantly reduces next-level applications and services development efforts. In this dissertation, we study the possible solutions to the problems above in an environment filled with IoT-ZigBee devices. We focus our work on the following research problems:

a. The representation of the IoT-ZigBee environment and its metadata and semantics in a suitable representation format can handle the complex relationships generated in such a densely populated environment.

b. We focus on efficiently collecting and processing streaming IoT-ZigBee data in real-time with minimum processing time and latency. We use a modular system that can be extended to support additional data types and devices with limited interventions.

c. They are focusing on the extraction of knowledge from the collected and processed IoT-ZigBee data. Raw data are of limited value to end-users in most cases. The real value lies in the conclusions that can be extracted from them.

Once we have achieved all objectives, we will be able to analyze IoT-ZigBee data in real-time for smart home with security generation offering the outcomes of our work to researchers that can then build upon our work and generate a brilliant and connected environment.

1.2. Research Contribution

We contribute to the field of data analytics for the Internet of Things in smart home environments and the post processing analysis of the IoT-ZigBee data by providing a better understanding of the physical surroundings with the help of solar system monitoring. We propose new and extensible methods to store the information of IoT-ZigBee installations and our findings allow for fast lookups across the whole set of IoT-ZigBee devices in the network. Processing the data generated in real-time becomes a more streamlined workflow, using a standardized methodology that allows for changing the data analysis from a single point. Data generated can be easily stored and retrieved from a specifically designed, reliable and efficient storage engine. Using the original and calculated data in a feedback loop also allows us to build additional processing and analysis layers that generate more data and information resulting in more accurate outcomes. The results of our work can be categorized in the following points:

i. The first goal of our work was to setup a base set of infrastructures for bringing the physical world with its digital representations for smart and secure home using ZigBee technology. We wanted to be able to understand what it is where and how interacts with its environment as well as what it observes in real time. Our work towards this direction was twofold:

ii. We have developed a graph-based schema to represent IoT-ZigBee installations together with their semantics and meta information. The schema is not based on the traditional relational databases but builds on the ideas of graph theory and utilizes a new database model, a graph database. Each entity that participates in real world interactions or can be observed by IoT-ZigBee device is represented as nodes of a graph, while the interactions themselves are the vertices of the graph.

iii. We therefore build a web of entities and relations that are easy to visualize and traverse to find answers to various queries that may arise inside a smart environment. Such queries can look for the causes of events observed (e.g., what caused the rise of luminosity in room), the available information for an area (e.g., what information is sensed for the building) or even the social interactions and connections of individuals (e.g., which people use the same appliances).

iv. For the analysis of the data we provide a template implementation for setting up a system that can receive, process, and analyze an unbounded number of input data streams with no impact on its operation using ZigBee with based K-Means algorithm. Such data streams can originate from sources that range from single smartphones to city wide sensor installations.

v. The system is also capable of handling data streams that provide unbalanced volumes of information with no performance drops. The calculation methodology itself is developed as to be easily customized and extended, based on the data types and the calculations required in every environment it is deployed to ZigBee. This processing engine is capable of identifying the data from each of the sensing devices routing them to the appropriate processor without any prior per-device configuration.

2. RELATED WORK

As we are now rapidly approaching 2030, it is as important as ever to address the skills that will enable all citizens to make informed and well-thought choices as mentioned in (Zulkipli et. al., 2017). Also, another ubiquitous fact should also come to mind: we cannot manage what we cannot measure. It is necessary to monitor the impact of our current behavior or the effect of potential behavioral changes in order to have a clearer picture with respect to e.g., our everyday smart and secure home. Furthermore, environmental education, as part of the broader issue of science, is an important component of the EU cultural heritage. In fact, EU considers environmental education one of the most prominent instrument to influence human behavior towards more environmentally sustainable patterns (Stoyanova et. al., 2020). Hence, associating environmental education and game-based learning will lead to students taking over a leading role in the educational process, setting questions, investigating the possible answers, and looking for alternative explanations to come up with a fitting model.

Several software products for monitoring sensor data exist. It is able to capture sensor data from proprietary data formats or protocols and visualize them. Regarding visualization, discusses the most common approaches with respect to feedback design in eco-conscious work for the past decade, from the perspective of both ICT and psychology, providing insights to their strengths and weaknesses. The researcher in (Servida and Casey 2019). presented an example of a typical engineering-focused approach, that utilizes several skeuomorphic metaphors to enable smart home feedback creation on smartphones. While such systems provide end-users and developers with powerful tools and front ends, we believe they should also be paired with multiple approaches, offering multiple possibilities to interact with the system. Examples of large-scale smart metering deployments that used IoT portals (for electricity and water) and other visualization tools to support the system and engage end-users to participate as mentioned in (Kruger and Venter, 2019). Their findings support the notion that the use of multiple approaches, with respect to visualization and feedback, serves such purposes well. We have followed a similar line of thought while implementing our own user interfaces and will continue to evolve our approach in future revisions of the system.

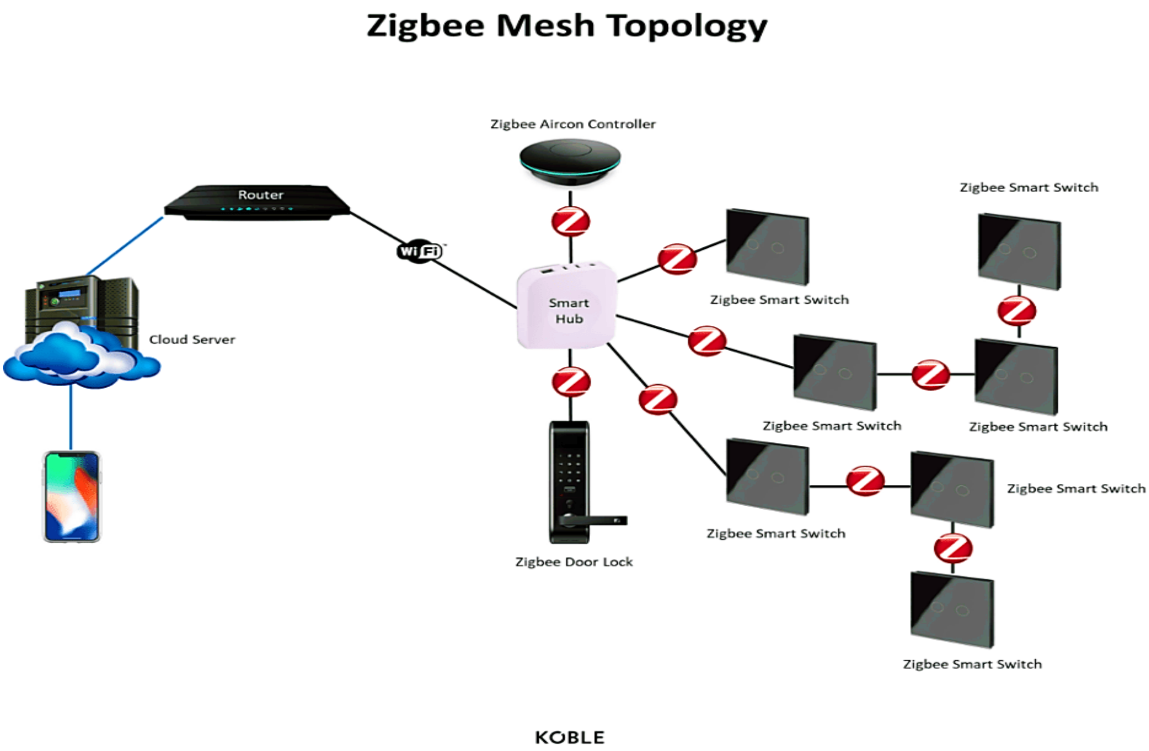

Figure 2. An advance home automation system schema developed for home using Zigbee-technology

In order to interact with an IoT installation, an appropriate representation in the digital world is required. Each installation comprises multiple elements that produce, consume or transfer data. Every element in this ecosystem potentially has multiple relations with other ones creating a complex schema that can be hard to understand or visualize as the scale of the installation increases as mentioned in (Yaqoob et. al., 2019). To overcome this problem and simplify our interaction with this digital image representation, formats more appropriate than the traditional relational databases need to be applied. At the same time, these formats need to be easy to understand and be used by non-tech-savvy users like artists or activists in their work with this, now accessible, living part of our environment.

In 2019 the number of Internet of Things (IoT) connected devices was 19.71 billion worldwide. By the end of 2020 it will hit 26.66 billion as mentioned in (Ahmed et. al., 2015). In 2025, the number of devices connected to the Internet will be 75.44 billion, the number will be risen 5 times in a decade. These smart devices are capable of generating data on its own based on behavior or expected result and share it over the Internet. This is what constitutes the concept of Internet of Things (IoT). IoT is a massive group of devices containing sensors or actuators connected together over wired or wireless networks as mentioned in (Bradac, et. al., 2015).

Besides all the benefits and positive characteristics that the wireless medium has brought to IoT, there are vulnerabilities. There are different forms of traffic anomalies and attacks. The most common security threats are intruders, which is generally referred, as hacker or cracker and the other is virus. This comes with the risk of personal data being broadcasted to the world and cause more cyber-attacks like phishing, denial-of-service (DoS), Probe, Remote to Local (R2L) attack etc. Since these attacks are not common so normal data flow has a pattern which can be more or less the same. However, when another entity will try to distort or change the data, the pattern will change, and this is where the Intrusion detection techniques comes into play in the context of internet and specifically in the context of Internet of Things as mentioned in (Collotta and Pau, 2015).

To monitor and detect anomalies or any suspicious behavior, intrusion detection system (IDS) and intrusion prevention system (IPS) are being used. As diverse environments and latest technologies are prone to be maliciously attacked, machine-learning (ML) algorithms are able to detect, analyze and classify intruders in network accurately and rapidly. Two types of intrusion detection system are there. They are Anomaly based detection and Signature based detection. Anomaly based detection methods detects the attacks that are unknown by observing the whole system and the entities in it such as traffic, objects etc. Anything other than normal behavior is identified as a potential attack. Conversely to detect specific patterns of the attacks that are known by investigating network data or traffic, signature-based detection methods are powerful as mentioned in (Pascual et. al., 2014).

ZigBee provides several day-to-day advantages not only for consumer's sector but for business sector as well as mentioned in (Baliyan et. al., 2015). This technology allows us to view real time information that was not available before. Business can improve their production efficiency by reducing material waste and unforeseen downtime. Since Artificial Intelligence is used, it saves considerable human effort and time. Sensors can be used to build infrastructure, detect damage to infrastructure. Automated traffic control can help reduce road congestion and gadgets or sensors can be implemented on the outside to recognize change in the environment and warn us for any impending natural calamity.

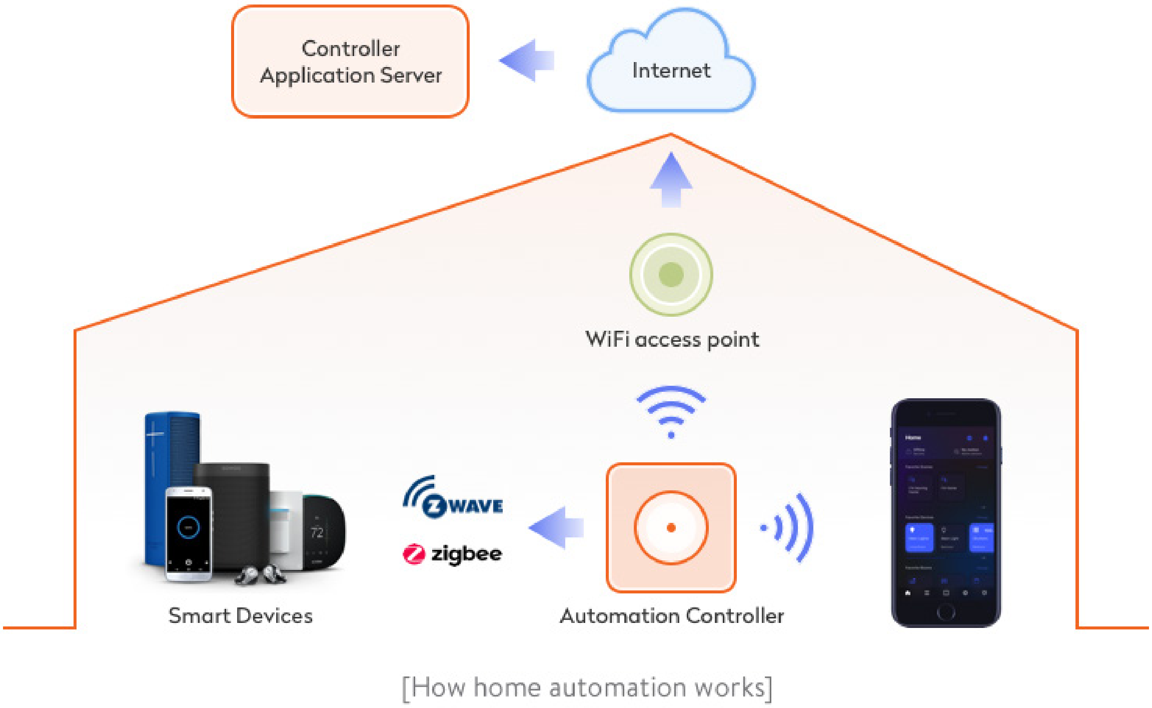

Figure 3. The ZigBee support with the automation controller described Source.

ZigBee makes our homes, offices, and vehicles smarter, more measurable and this in turn makes our life more pleasing. Smart speakers like Amazon's echo or Google home makes it very easy to set timers, get information on the go or be it to play music. Through home security system, we can monitor our home from other places remotely and can see what is happening on the inside and outside or talk to visitors as mentioned in (Marino et. al., 2016). Smart lighting systems can reduce unnecessary operation if there is no one in the house , and smart air conditioners will turn on when it knows you are on your way to your home. This makes them prone to security risks like DDoS attacks, one of the most common issues, which happens by setting the device's password to its default password which are easily crack able by hackers. Ransomware and Malware both relies on encryption to lock out the user's devices completely and steals the user's data. Privacy remains among the largest issues of ZigBee as data is being transmitted, stored, and processed and being harassed by large companies. Internet of things disconnection happens in smart homes due to many reasons. It could be because of a high growth of ZigBee in the world. Other reasons could be due to the movement of nodes, which could be a random walk to single or both nodes of each link or could be periodic due to satellites and planets movements on orbits. It is good to mention also that some cut offs occur due to nodes periodic power saving, like low power nodes in sensors networks as mentioned in (Niaz et. al., 2017).

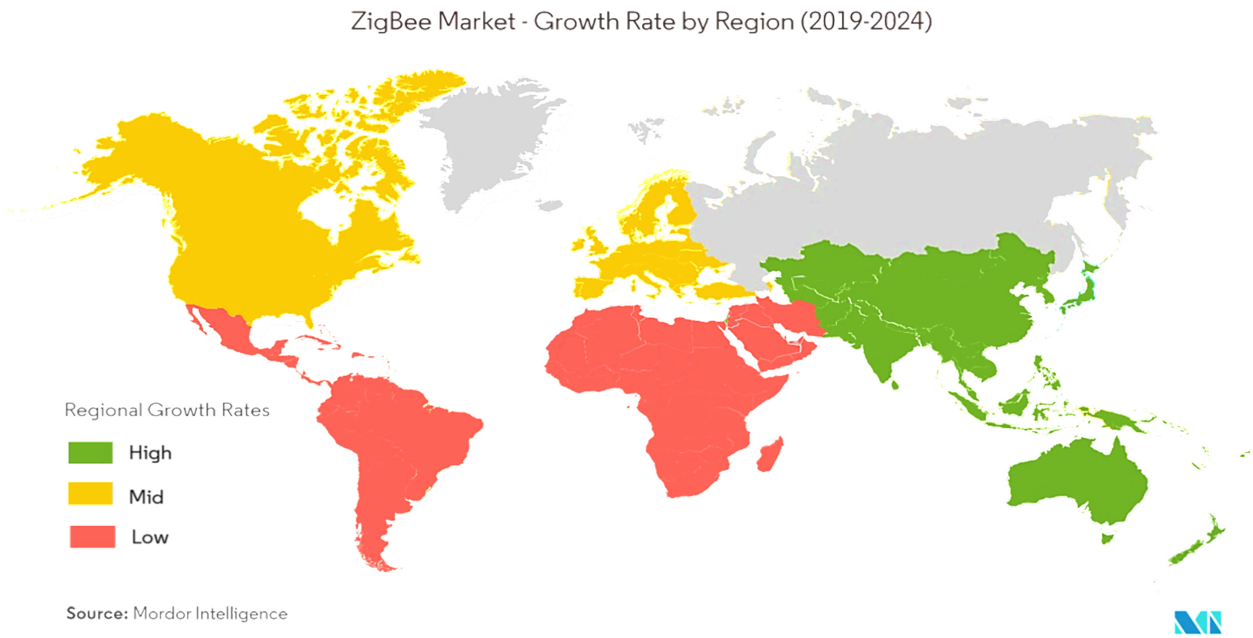

Figure 4. Interest over time according to ZigBee predicted growth from 2019 to 2024 for terms Smart device and Internet of Things

3. METHODOLOGY

The research is organized in various sections starting with a description and the philosophy of the research from which the electricity consumption data and house information is taken and used as input for the study; a remark is done in how consumer behavior can contribute to foster energy efficiency; then the objectives and goals are defined. Moving to the more technical part, the procedure of cleaning the raw data is explained to have the processed data in order to start the data analysis sought. First, an exploration and visualization of the data was carried out to extract the first outcomes from the consumer's load profiles. The core of this research focuses on the smart home network segmentation according to their electrical load profiles, various clustering methods are used and compared in order to find the most adequate consumer segmentation. Then, the house features and household characteristics are considered to find whether exists any relation between customer's segments and their house properties or not. The data generated has increased exponentially this last decade, this data could contain valuable information, but it needs to be explored. The electricity market is also living a data boom, for instance the electricity meter reading has moved from one read every month to host smart meters able to read electricity consumption every smart home network; this implies that each consumer will have around 100 measurements per year. Store and manage this data are already a challenge itself, that is the reason why smart home networks and computer science have increased their presence in the energy sector; but also, the possibility to add value to this massive amount of data applying data mining and analytics to extract hidden information or at least be able to separate the useful one to the less interesting. Big data analytic tools are essential to add value to all this data, specifically in the energy sector could add value in terms of energy balance, energy efficiency or energy prediction; that could be profitable for both the consumer and the utility.

3.1. Dataset Description

A dataset is composed of rows and columns, where, usually, each row represents an observation and each column a different variable. The dataset is named as "Smart Home Dataset with weather Information" which is freely available on KAGGLE. This data can be either qualitative (categorical) such as gender, country, city; or quantitative (numeric) such as electricity consumption, height, price, distance. Again, and as said before depending on the type of data the approach used to the analysis may vary, because is no treated the same way numeric and categorical data. The size of this dataset is relatively manageable due to mainly three facts. The dataset can be downloaded from the link: https://www.kaggle.com/taranvee/smart-home-dataset-with-weather-information.

i. The number of sub-meters installed are a small amount accounting for different sub-smart home networks.

ii. The measurements period is in most of the cases around year length.

iii. The data was aggregated hourly, fact that reduced the final number of observations and therefore the dataset's size consisting of 124.89 MB.

iv. Electricity consumption data for smart home networks with weather information.

v. Households and householder's information based on smart home network usage.

3.2. Communication Via Smart Home Network

Communication via smart home network solves many problems that directly affected the data quality, as if the smart home network is switched on, communication can be stablished interrupting the sending and reception of data. Then, the data for this time period is represented by zeros as it was no consumption; thus, altering the dataset values which need to be eliminated in order to perform an accurate data analysis.

From the database server the Software as a Service (SaaS) company, in this case, access to the data stored and proceed to upload it into the web-app platform for the visualization of each user. Finally, the data used for the current analysis was obtained through an Application Programming Interface (API) configured by this research in order to:

vi. First, aggregate all the data hourly. The sub metering equipment is able to measure the consumption in high-resolution frequencies, each 5-15 minutes according to the device. For the aim of this study, the hourly aggregation is enough and a good starting point if a further deeper study is sought in the future.

vii. Second, all the household's consumption data are in the same file, fact that facilitates the analysis procedure as reading one file is enough to analyses the data.

viii. Hence, by using the "python programming" that permits the access to the database server a ".csv" file is downloaded, this file is the input data regarding electrical consumption that will be used for the data analysis. This data is considered time series numerical as it records the electrical consumption and time of it.

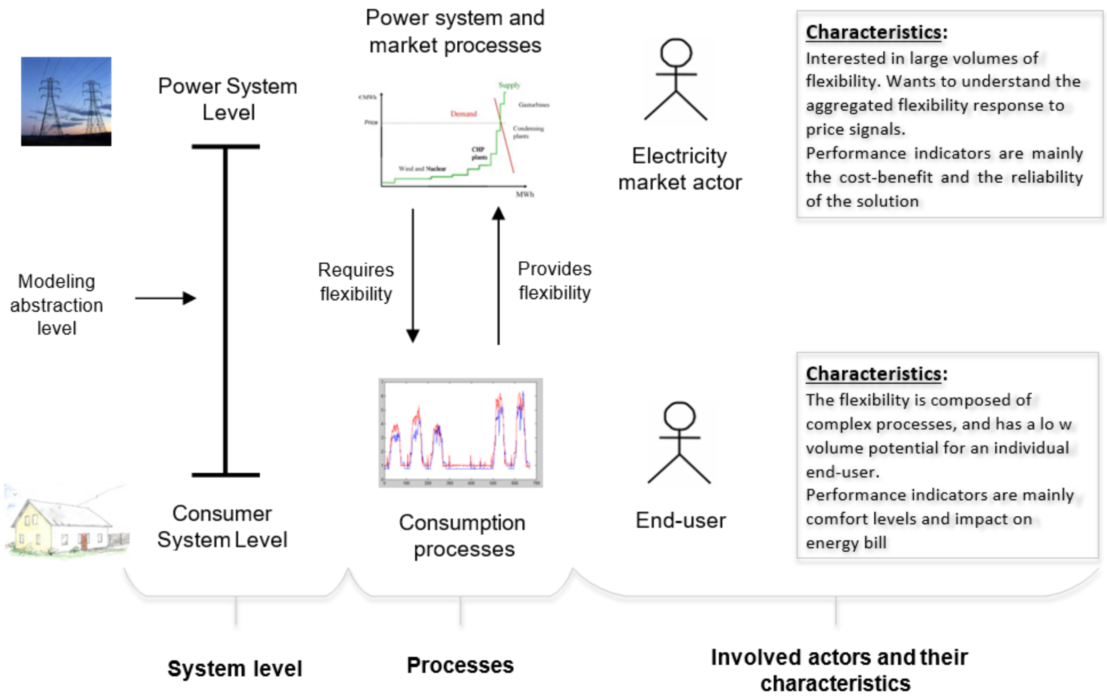

Figure 5. Diagram that describes the functions of power system for maintaining the secure home network.

The ".csv" files have the characteristics that are easy to read by software prepared to conduct a data analysis, for instance "MS Excel", "Python" among others. "Python" is the programming language used to carry out the current analysis.

3.3. Clustering Algorithm for Prediction

The K-means solution adopted was to cluster the data in 7 groups by using the python programming language, but as said before the applying the algorithms could also be chosen. Due to the fact at this stage is not possible to stablish big differences among spectrum allocation of smart home network. The number of members assigned to each cluster, range from the minimum different number of clusters for smart and secure home data; so, a distributed user's allocation is also achieved; avoiding clusters with 1 to 5 members which could mislead the analysis. Table 3.1, as expected, the clusters with large number of networks present a higher within cluster distance.

Table 1. Number of members per each cluster defined by K-Means algorithm using ZigBee protocol.

K-Means Algorithm on smart and secure home Data using ZigBee Protocol							
Cluster Number	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Number of members	9	17	8	8	37	8	35

Cluster 1: there are two peaks, morning, and evening both of them represents around 7% of the energy usage. It is similar to the cluster 6, although this cluster one presents a higher proportional value between peaks that ranges from 3-5%, it could mean that some activity is done for smart home networks.

Cluster 2: Presents a not sharp morning peak and a low consumption percentage during day of around 3% similar to the night hours' proportion; this means that there is no activity at home during the day.
Cluster 3: Presents a high morning peak, around 9% of the daily energy use is consumed in that peak that is around 8:00h. Then, there is no activity as the proportion of energy use in the afternoon decreases down to 3%, similar at the night percentage, and there is a small evening peak of around 5% at the dinner time.

Cluster 4: It presents two peaks prolonged in the morning (9:00) and in the afternoon (19:00), following the schedule of a small business or local, also the night consumption proportion is so low. So, it seems it is not a residential consumer.

Cluster 5: This cluster presents a flat profile, with the higher values during the day at lunch and dinner time. It could mean that people stay home during the day.

Cluster 6: Presents two sharp peaks, one in the morning and other in the evening, that reach almost 7% and 10% respectively. Then the energy usage during the rest of the day and during the night is so low, constant at 2.5%, so no activity during these periods; it means that the households are not home during the day; could correspond to working people with children.

Cluster 7: Presents a significant evening peak around 21:00h of almost 9%. There is also a small morning peak, almost no noticeable which does not reach the 5% of energy usage. During the days it also presents a continuous consumption of around 4- 5% until the evening peak.

4. RESULTS

The electrical consumption is analyzed in different levels of detail. The followed procedure starts from analyzing the whole gross data and gradually moving towards a more detailed data analysis by segmenting and grouping the data for a more accurate study. Two kinds of values were used to plot the electrical consumption; the absolute units of power consumption [kWh] and the proportional hourly energy usage in percentages (%) were used. When analyzing each consumer individually the important values are the absolute units in terms of kWh; but to fairly compare various consumers' profiles is necessary to create a common normalized framework where to contrast the curves' shape; this can be done with the hourly percentages of energy consumption. As said before, depending on the analysis' goal one parameter or another would be used with ZigBee protocol. So, although the possible approaches to the consumption data an endless, below different representation ways are shown that would be useful to know and better understand the consumer. This section uses various consumers for the different representations, in order to see variety of load profiles and also test the validity of the illustrations. Remark that, the scope of analyzing the features for each cluster would have been more adequate if a larger sample with better data quality and more specific features were available. So, the current process could not add any value to the research; but it is a procedure that could be followed when a larger dataset would be available.

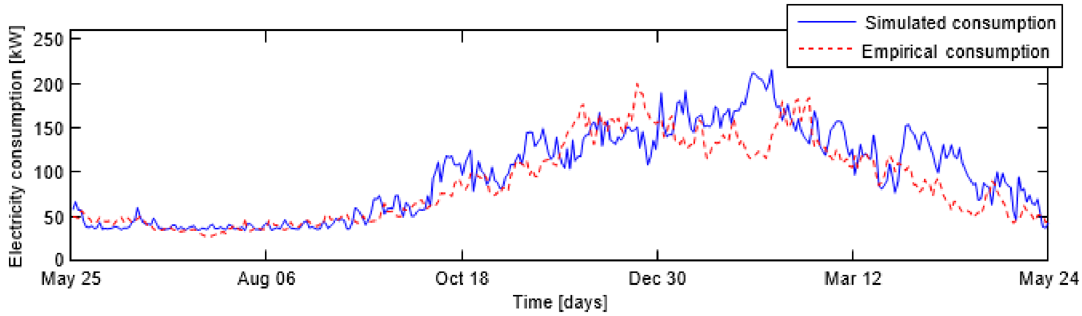

Figure 6. The smart home energy consumption expressed in absolute time in terms of consumption units (kWh), in graph bars for simulated and empirical consumption.

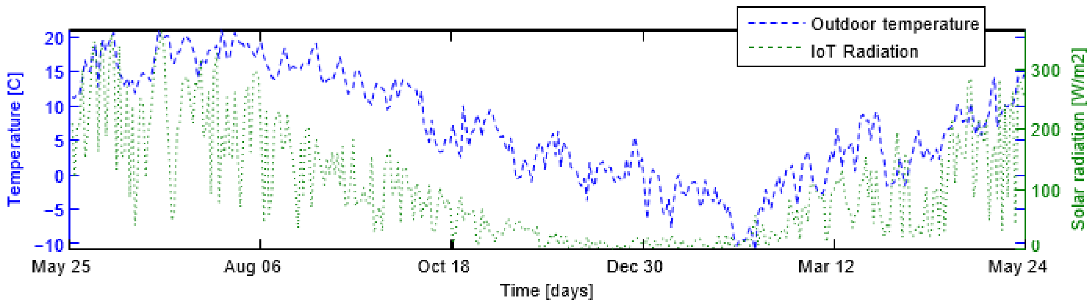

Figure 7. The smart home energy consumption expressed in absolute time in terms of consumption units (kWh), in graph bars for outdoor temperature and IoT based ZigBee radiation.

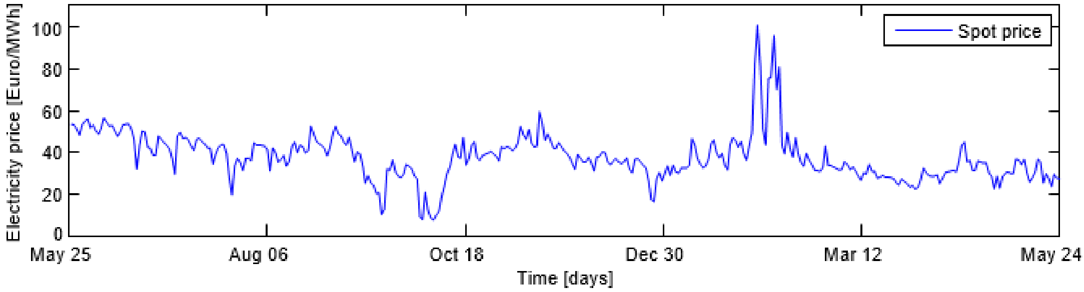

Figure 8. The smart home energy consumption expressed in absolute time in terms of consumption units (kWh), in graph bars for spot price including security.

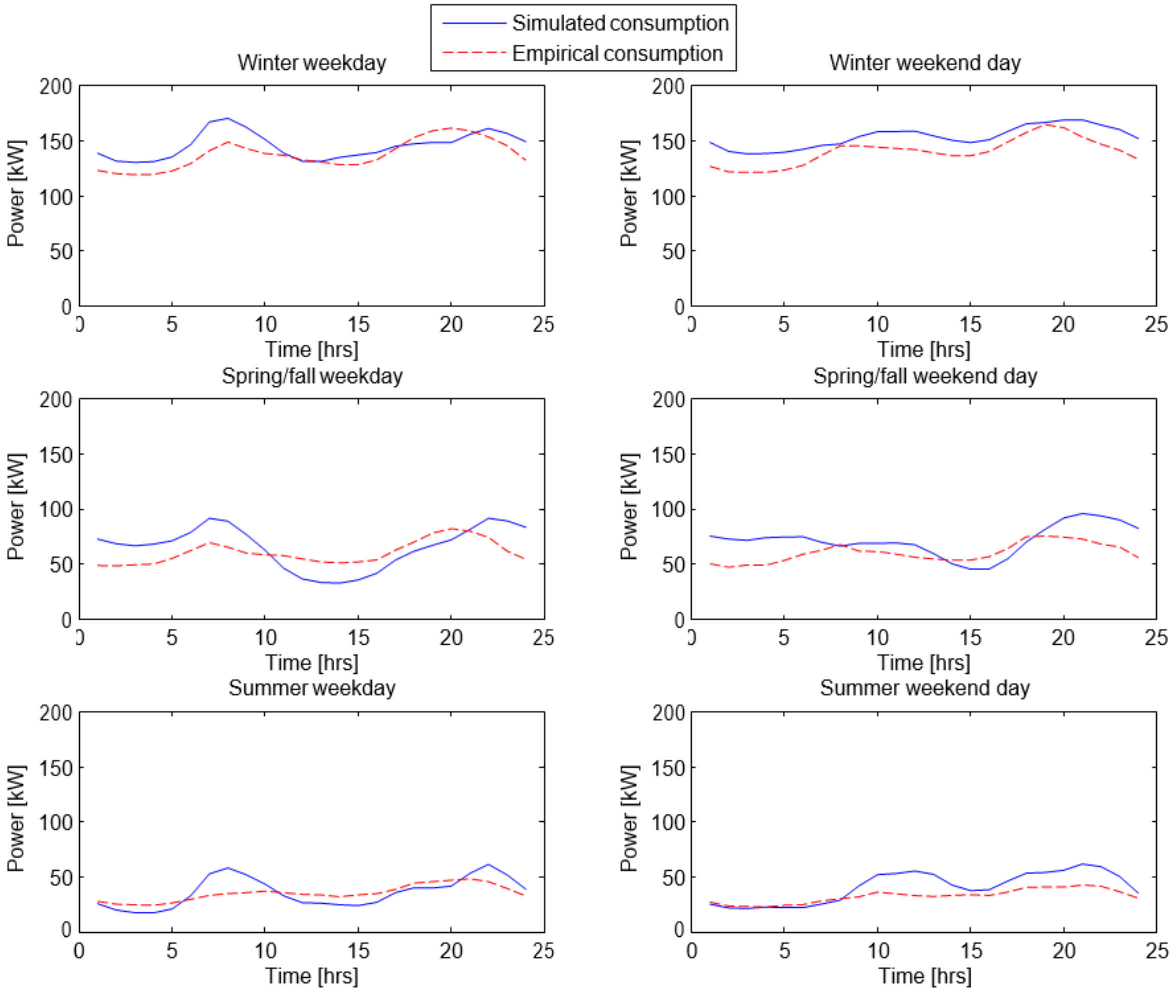

Figure 9. The power consumption plot per season of winter, spring and summer expressed in absolute time (hrs) as well as usage, in graph bars for simulated and empirical consumption in smart homes.

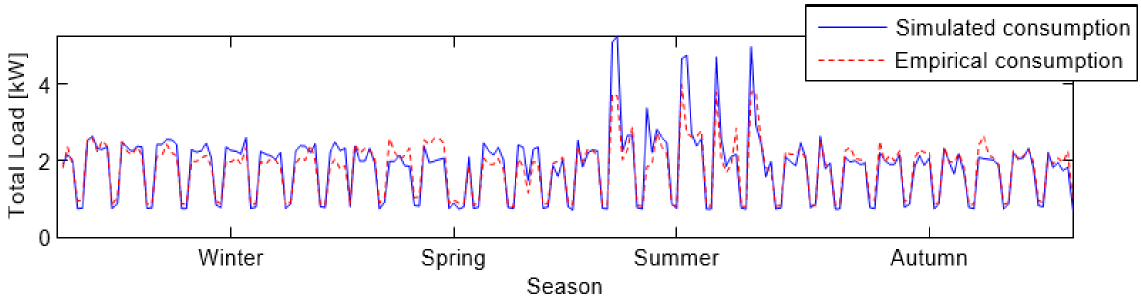

Figure 10. The prediction of energy consumption for all four season in a year with respect to total load (kWh), in graph bars for simulated and empirical consumption in smart homes.

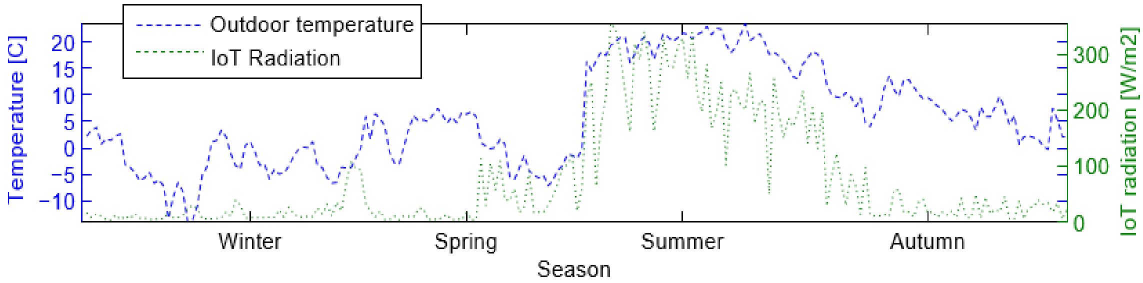

Figure 11. The prediction of energy consumption for all four seasons in a year with respect to home temperature (C°), in graph bars for outdoor temperature and IoT based ZigBee radiation.

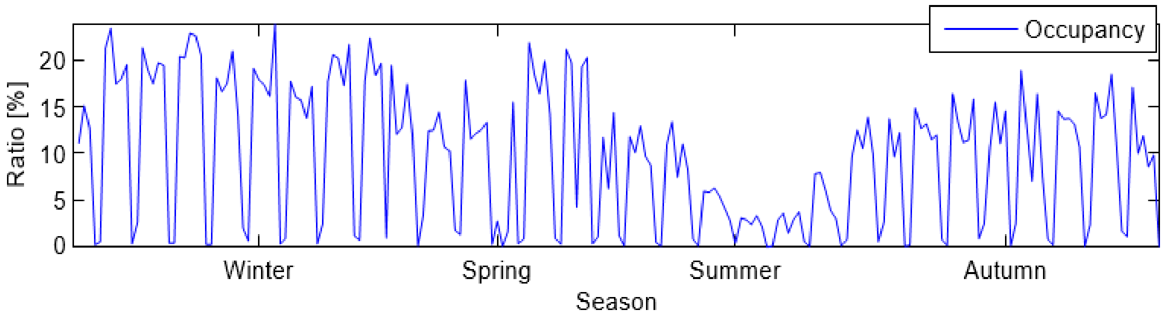

Figure 12. The prediction of energy consumption for all four seasons in a year with respect to usage ration (%), in graph bars for consumer occupancy.

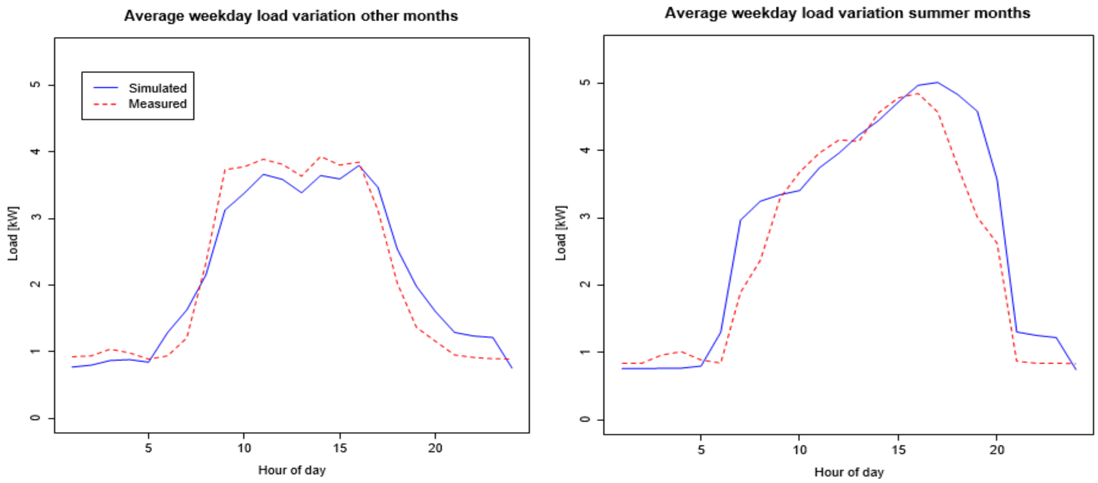

Figure 13. The average weekday load variation for summer and other months with respect to hour of day where curves expressed in hourly absolute load consumption units (kWh) for simulated and empirical consumption using K-Means.

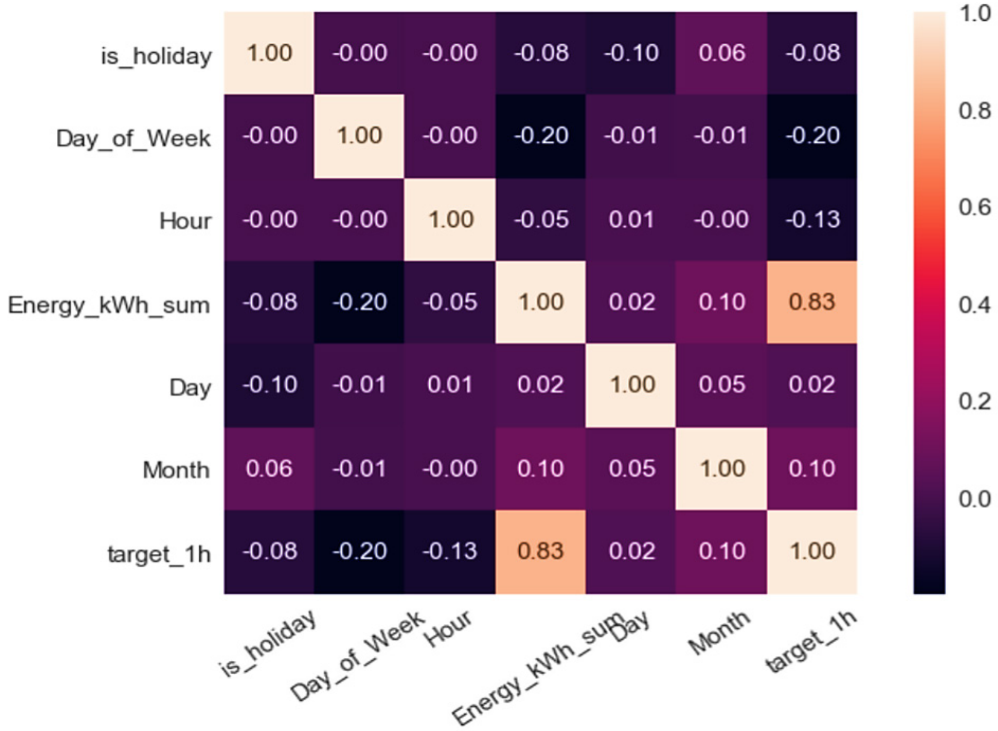

Figure 14. The confusion matrix for average energy usage per week with minimum correlation coefficient for multiple households.

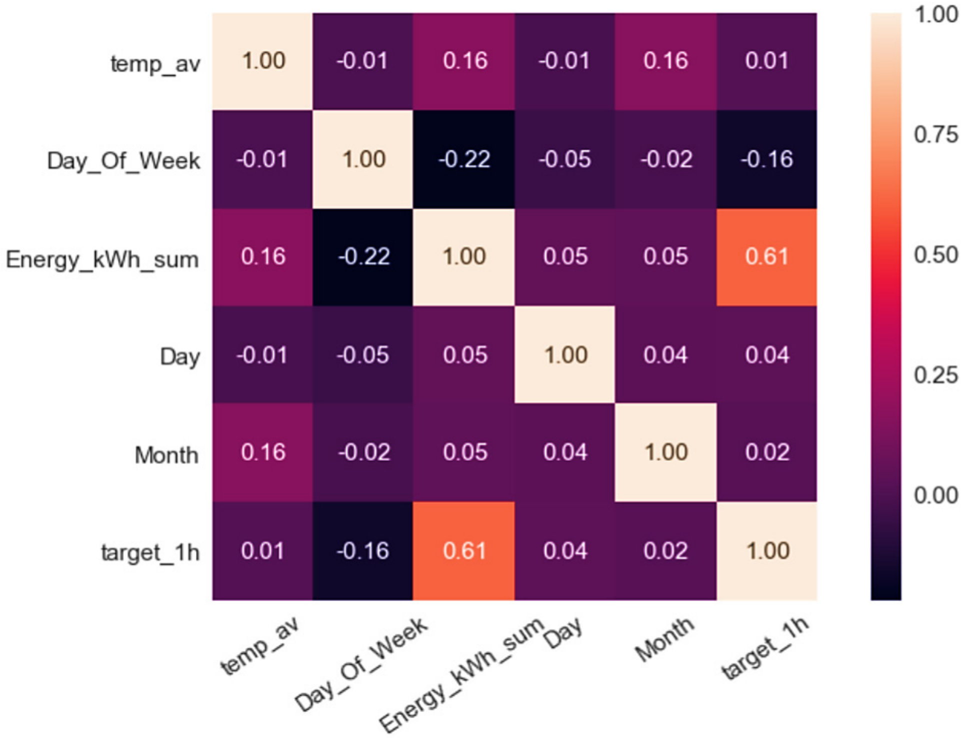

Figure 15. The confusion matrix for average energy usage per week with minimum correlation coefficient for single households.

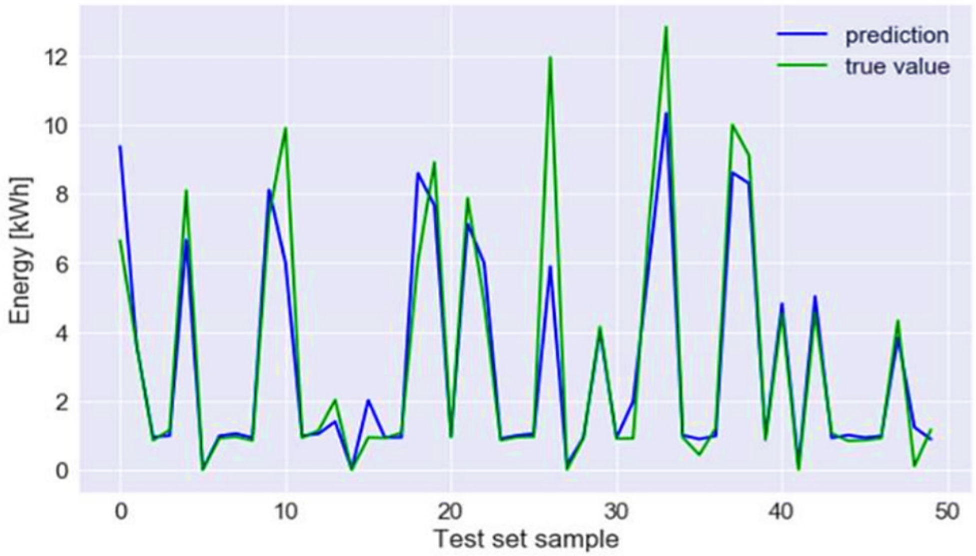

Figure 16. The forecast of energy usage in terms of true value and prediction on 50 household test samples.

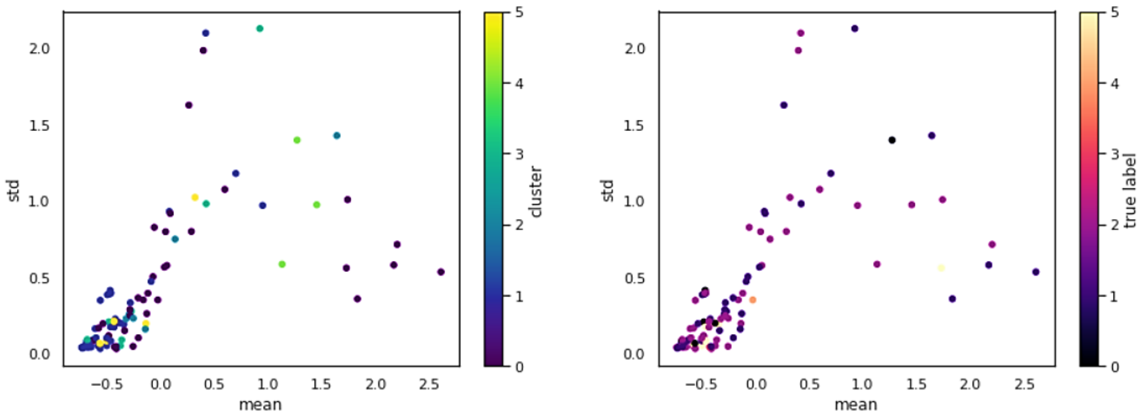

Figure 17. The Clustering the Energy Consumption with its rela

4. DISCUSSION

Energy management tools are not extended within the residential sector, since the energy consumed is so small compared to business and industries, the savings obtained do not pay-off the investment. However, there is room for providing such tools to residential consumers, for instance by creating a community as ZigBee protocol are doing with the collaboration of public administrations and electricity utilities. This research is an example of a public administration pioneer initiative to engage the consumers and foster the energy efficiency among them, aiming to provide energy knowledge, understanding and guidance to the user in order to reduce its consumption on average up to 10%. Nonetheless, it was seen that providing the technical consumption data in kWh to the householders has limited influence and effect, so is necessary to go beyond and translate the technical information into call-to-action measures, guiding the user to smart energy choices (like a GPS), have a better response from the householder. Also, the need of sub-metering devices to obtain the consumption data limits the scalability, as the cost associated is significant. To be able to reproduce this kind of research in a large-scale the access to the data from the utilities smart meters is necessary. So, the combination of the smart meter deployment and the big data analytics are called to play an important on the energy sector. The smart meters generate large amounts of raw data that need to be managed and, once analyzed can be converted into useful information that benefits both the utility and the consumer, as aim to improve the customer engagement and the quality of the service. Creating new business opportunities, mainly related to data science and data analysis in response to the market needs.

Table 2. The comparison of proposed method of implementation with existing technique.

ARTICLE	TECHNIQUE	ACCURACY
Peng et al (2019)	Support Vector Machine (SVM)	89.97%
Jaihar et al (2020)	Convolutional Neural Network (CNN)	96.73%
Proposed	ZigBee based IoT Protocol with K-Means	98.46%

Most of the home appliances features are common to each cluster, with energy usage data in the broadband network, and not available data for the domestic hot water system. Also, it does not exist any significant difference among the house properties and the sociological features in the different clusters, in order to be able to affirm that one or some specific features are the cause of the load profile pattern. Remark that, the scope of analyzing the features for each cluster would have been more adequate if a larger sample with better data quality in ZigBee protocol and more specific features were available in the result section of this research. So, the current process could not add any value to the research; but it is a procedure that could be followed when a larger dataset would be available.

5. CONCLUSION

Smart home energy consumption prediction is the key to perform a proper analysis and extract reliable conclusions from it. In this study, the dataset was the optimal in terms of quantitatively and qualitatively. This Kaggle dataset utilized well as instance counting for thousands of users, would have been more adequate to the purpose of the study than the small sample studied of only users. In addition to that, the electrical consumption data from the sub-metering equipment presented some issues that difficult the analysis at some point and demanded a pre-clustering stage for treating and cleaning the data. As the communication between the sub-metering device and the server is done through the broadband network of the consumer, when it is switched-off the consumption data is lost and counted as high spectrum allocation like if it had been no consumption. So, when the possibility to access to the hourly smart meter consumption data becomes a reality, the current study purposes will find a suitable framework since the numbers of users could be much higher and the data quality should be almost perfect as are the measures from the electricity utility to bill their customers. The methodology used in this research is simplified compared to the procedures of the referenced papers, as larger datasets and more complex parameters are used. normalize the electricity consumption considering the weather and seasonal effects, although for the present study there was no temperature data available. The K-Means clustering technique was tested and compared, giving best prediction of accuracy 98.46% for all energy load profiles with high spectrum of 100GHz. The solution from the ZigBee Protocol based K-Means clustering is the one that better adapts to the segmentation sought, which is used as the base of the post-clustering stage to obtain the final energy consumption segmentation. This normalization is advisable as it mitigates the electricity consumption correlated to the temperature; however, this is more appropriate for aggregated daily consumption. But for instance, the electricity consumption is not normalized with the temperature. Most of these methodologies use the absolute values in 100 kWh, as they were more focused on identify the users with higher energy savings potential. But for the segmentation purpose, the decision to use percentages of energy usage fulfils the expectations sought.

6. REFERENCES

Ahmed, M.A., Y.C. Kang, and Y.-C. Kim. 2015. Communication Network Architectures for Smart-House with Renewable Energy Resources. Energies, 8, 8716–8735.

Aslan, J., K. Mayers, J.G. Koomey, and C. France. 2017. Electricity Intensity of Internet Data Transmission: Untangling the Estimates. J. Ind. Ecole, p 1–14

Baliyan, A., K. Gaurav, and S.K. Mishra. 2015. A Review of Short-Term Load Forecasting using Artificial Neural Network Models. Procedia Computer. Sci., 48, 121–125.

Becirovic, S., and S. Mrdovic. 2019. Manual IoT Forensics of a Samsung Gear S3 Frontier Smartwatch. In Proceedings of the 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, pp. 1–5.

Björnson, E., L. Sanguinetti, J. Hoydis, and M. Debbah. 2015. Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the Answer? IEEE Trans. Wire. Commun., 14, 3059–3075.

Bouktif, S., A. Fiaz, A. Ouni, and M. Serhani. 2018. Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches. Energies, 11, 1636

Bradac, Z., V. Kaczmarczyk, and P. Fiedler. 2015. Optimal Scheduling of Domestic Appliances via MILP. Energies, 8, 217–232.

Collotta, M., and G.A. Pau. 2015. Novel Energy Management Approach for Smart Homes Using Bluetooth Low Energy. IEEE J. Sel. Areas Commun. 33, 2988–2996.

Ge, X., J. Yang, H. Gharavi, and Y. Sun. 2017. Energy Efficiency Challenges of 5G Small Cell Networks. IEEE Commun. Mag, 55, 184–191.

Jaihar, J, L. Neehal, S.V. Patel, V. Gautam, and K.P. Upla. 2020. Smart Home Automation Using Machine Learning Algorithms. 1-4. 10.1109/INCET49848.2020.9154007.

Kruger, J.-L., and H. Venter. 2019. Requirements for IoT Forensics. In Proceedings of the 2019 Conference on Next Generation Computing Applications (Next Comp), Mauritius, pp. 1–7.

Marino, D., K. Amarasinghe, and M. Manic. 2016. Building Energy Load Forecasting using Deep Neural Networks. In Proceedings of the IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy.

Niaz, M.T., F. Imdad, and H.S. Kim. 2017. Power Consumption Efficiency Evaluation of Multi-User Full-Duplex Visible Light Communication Systems for Smart Home Technologies. Energies 10, 254.

Pascual, J., P. Sanchis, and L. Marroyo. 2014. Implementation and Control of a Residential Electrothermal Microgrid Based on Renewable Energies, a Hybrid Storage System and Demand Side Management. Energies, 210–237.

Peng, Y., J. Peng, J. Li, and L. Yu. 2019. Smart Home System Based on Deep Learning Algorithm. Journal of Physics: Conference Series. 1187. 032086. 10.1088/1742-6596/1187/3/032086.

Stoyanova, M., Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E.K. Markakis. 2020. A Survey on the Internet of Things (IoT) Forensics: Challenges, Approaches and Open Issues. IEEE Commun. Surv. Tutor, 22, 1191–1221.

Servida, F., and E. Casey. 2019. IoT forensic challenges and opportunities for digital traces. Digit. Investig, 28, 22–29.

Wu, G., C. Yang, S. Li, G.Y. Li. 2015. Recent advances in "energy-efficient networks and their application in 5G systems. IEEE Wire. Commun., 22, 145–151.

Yaqoob, I., I.A.T. Hashem, A. Ahmed, S.A. Kazmi, and C.S. Hong. 2019. Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges. Future Gener. Computer. Syst., 92, 265–275.

Zulkipli, N.H., A. Alenezi, and G.B. Wills. 2017. IoT Forensic: Bridging the Challenges in Digital Forensic and the Internet of Things. In Proceedings of the 2nd International Conference on Internet of Things, Big Data and Security (IoTBDS 2017), Porto, Portugal; pp. 315–324