

# Network Security Monitoring of Smart Home System Using Machine Learning and Data Mining

Salwan Mahdi Alshabeeb<sup>1</sup>,

Adil Deniz Duru<sup>2</sup>,

Abdullahi Abdu Ibrahim<sup>3</sup>

Information Technologies Department 1,

Physical Education and Sports, Marmara University 2

Software Engineering Department 3

1,3Department Of Electrical and Computer Engineering,

Altinbas University, Institute of Science, Istanbul, Turkey

Kokyalshekh14@gmail.com 1, Abdullah.ibrahim@altinbas.edu.tr 3

## Abstract

The goal of this hypothesis is to make a trap of tasks insightful home with the daylight-based power worldwide situating structure that can be presented over existing WiFi associations, going without moving cycles and allowing fashioners to get to or kill equipment from the system buyer 's needs, all while including a virtual article for robotization and energy interest. With the amazing advancement of Internet-of-Things (Machine learning) use and devices, as well as clients' outrageous interest in hoarding these devices to additionally foster comfort and redesign family tasks, it is essential to manufacture a strong and fruitful far off frameworks organization system that lets sharp homes to affix these contraptions and sort out their nuances in a strong way, with low set - up costs using daylight-based energy. So, the goal of this work is to create a response for an insightful home system that uses just WIFI networks for correspondence, avoiding intrusive foundations (no extra actual wires required, essentially viable python programming considering on state-of-the-art replicating), and allows to interact and control various contraptions, from any carrier, from a separation from a PDA, tablet or another device with an affiliation Internet. A muddled and estimated structure has been made, using standard advances and shows and involving a central server (clever home server), which gives a lot of organizations to control and manage the system; an associate informational collection to restrict the prerequisite for information contraptions to store; a MQTT delegate licenses devices to convey using this show, MQTT ultimately devices with unimportant necessities, they just ought to have the choice to connect with a WiFi association. The system has no limitation on the number of related devices as its designing has been arranged and executed to think about level flexibility and high availability to



supervise progressively more power consuming devices. sun controlled energy with a low unit use obligation.

**Keywords:** Internet of Things, Machine learning, MQTT, Intelligent Home, Home Automation, WiFi Network, Interoperability, Solar, Monitoring.

## Introduction

In the Internet of Things, physical items are implanted with electronics, software, sensors, actuators, as well as connectivity that allow them to share data with one another and with the rest of the world using their unique identities, as well as with the rest of the world via the Internet [1]. It is possible that the physical and digital worlds may eventually merge in ways that will benefit people across the world if we have this type of infrastructure. From smart grids to smart greenhouses to smart transportation, all these applications have profited from their implementation. A few examples of products that have proven to be successful in the industry include personal fitness trackers, smart lighting, smart greenhouse solutions, and other similar products [2].



Figure 1. A smart and connected home [2].

Although there has been a tremendous increase in the number of IoT applications, the value of the data produced through IoT systems has yet to be determined and confirmed. Since an individual sensor and wearable device may generate a large amount of data that is beyond the device's ability to handle, the data is often transferred to the data processing application where it can be viewed and used by others.

[3] A more favorable use has been developed. As these sensors grow increasingly intertwined into our real activities, the implications of storing such a large amount of data in the cloud may become increasingly significant. For example, when compared to information about a person's lifestyle and nutrition ( for example, information about a person's daily step count and journeys as latitude-longitude coordinates), information about a person's daily step count and journeys as latitude-longitude coordinates are relevant [4]. Therefore, a comprehensive strategy for data management is currently under investigation, both in academic and business settings alike. In the hope of attaining a global standard, numerous interfaces of transmitting as well as representing such data formats have attempted to do so, but none has taken the initiative to do so to yet.

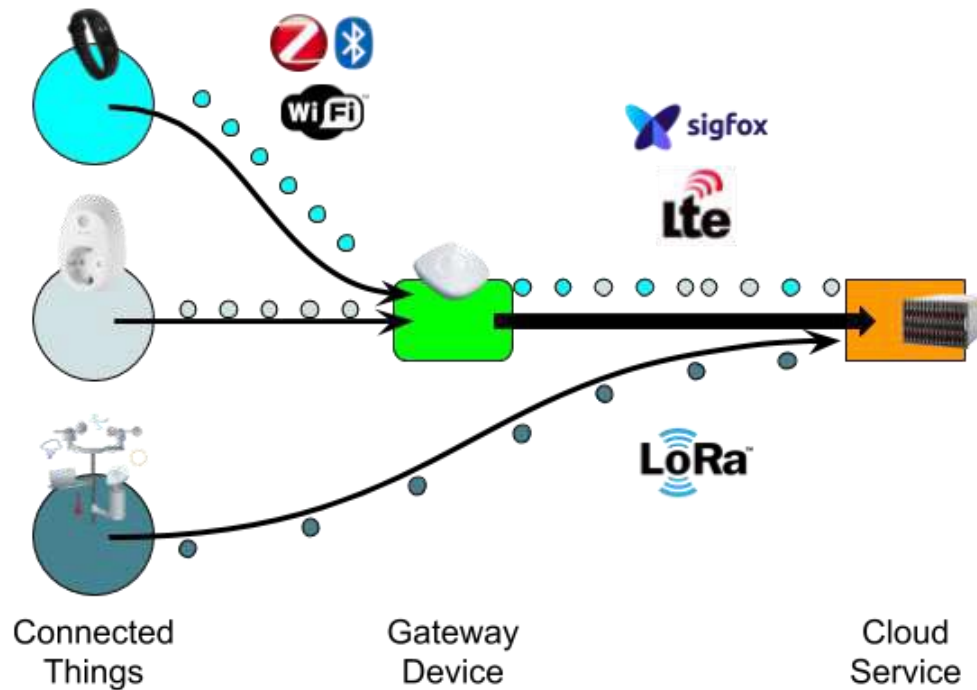


Figure 2. Data Streams of the Internet of Things [4].

Another distinguishing element of the Internet of Things (IoT) sector is the large number of communication mediums, protocols, other data representation and sharing formats that are being employed in the industry. Devices connected to the Internet of Things (IoT) can communicate using cloud services using a number of technologies and protocols, as illustrated in Figure 1.2 and explained in [5]. There are a variety of protocols to choose from, based on the throughput, range, and power consumption needs of the application. Devices such as smart greenhouses and wearables interface with local gateways via low-power communication protocols such as Zigbee[6], Bluetooth LE[7], or WiFi [7], which are all available. The usage of battery-powered devices like Zigbee and Bluetooth LE is preferable because these technologies are more versatile (e.g., wearable fitness trackers or beacons). WiFi is the superior choice for devices that require real-time connectivity and the ability to exchange large volumes of data [8]. To accommodate the higher power requirements of WiFi, these devices are generally typically connected into a standard wall socket. This is since using batteries for an extended period is impracticable. There are a variety of technologies that gateways, or smart devices can employ to communicate with the Internet, including DSL lines, LTE [9], LoRa [10], and Sigfox [10]. In order to compensate for this, DSL as well as LTE are more expensive to establish and maintain than LoRa and Sigfox networks. They have a higher throughput, however, because they require on-premises infrastructures, as well as ongoing maintenance and recurrent expenses, which reduces their cost. These two choices are easier and less expensive because they do not require a monthly subscription charge for Sigfox and little (or no) costs for LoRa, but their throughput is significantly lower than that of more expensive solutions such as ZigBee and ZigBee Plus. The sensing units of a device can only sample at a specific rate, which varies based on the communication module used in each of the instances above. The most common guideline currently is to sample and send data as soon

as possible while still maintaining a fair amount of battery power, which is followed by most applications and devices. As a precaution, no scheduled activities will be missed because of devices sleeping in order to conserve power, as indicated in [11]. To draw meaningful conclusions from an ever-increasing volume of data, it is necessary to evaluate that data.

## 1.1. PROBLEM STATEMENT

We focus our work in the following research problems:

2. The portrayal of the Machine learning climate along with its metadata and semantics in a reasonable portrayal design fit for taking care of the perplexing connections that are produced in a thickly populated climate.
3. The proficient assortment and handling of streaming Machine learning information, continuously with the base handling time and idleness, utilizing a measured framework that can be stretched out to help extra information types and gadgets with restricted mediations.
4. The extraction of information from the gathered and handled Machine learning information. Crude information is of restricted worth to end clients. The genuine worth lies on the ends that can be separated from them.

Once we have achieved all objectives, we will be able to analyze machine learning data in real time for smart home with solar power generation offering the outcomes of our work to researchers that can then build upon our work and generate a truly intelligent and connected environment.

## 4.1. RESEARCH CONTRIBUTIONS

The results of our work can be categorized in the following points:

- A. The primary purpose of our work was to establish a foundation of infrastructures for connecting the physical and digital worlds. We needed to know where it was and how it interacted with its surroundings, as well as what it was seeing in real time. Our efforts in this direction were threefold.
- B. To represent Machine Learning installations, as well as their semantics and meta data, we designed a graph-based architecture. The schema is based on graph theory and uses a new database paradigm called a graph database, rather than typical relational databases. Each entity that engages in real-world interactions or may be monitored by a Machine Learning device is represented as a node in a graph, with the interactions themselves serving as the graph's vertices.
- C. As a result, we create a network of things and relations that is simple to perceive and navigate to obtain answers to a variety of questions that may occur in a smart environment. Such queries can look for the reasons of observed occurrences (e.g., what caused the spike in room luminance), the accessible information for an area (e.g., what information is sensed for the building), or even individual social interactions and linkages (e.g., which people use the same appliances).
- D. For data analysis, we propose a template implementation for putting up a system that can accept, process, and analyze an infinite amount of input data streams while being operational. These data streams can come from everywhere, from a single smartphone to city-wide sensor systems.
- E. We also tested the processing engine for more than three months in real-world situations, obtaining sub-millisecond processing times each measurement. To demonstrate that our system is future-proof and scalable, we loaded it with data quantities that greatly outnumber the data collecting rates of our real-world installation.

## 5. LITERATURE REVIEW

To gain a better understanding of how to collect data from the more distributed and uncontrolled IoT installations, we also developed a method that can collect data utilizing cellphones in smart cities carried by volunteers all through crowd-sourcing campaigns in order to gain a better understanding of how to collect data from more distributed and uncontrolled IoT installations. This strategy is designed to supplement smart city deployments with low-cost, mobile, as well as volunteer-powered infrastructures in order to maximize the benefits of smart city technology. These solutions, which may be implemented at a low cost by city officials as a new service, will provide them with useful information about the present state of affairs in the area. More than 50 individuals have participated in experimentation activities in cities around Europe, including this smartphone application, over the course of the preceding two years [13].

A framework for categorizing information from all of the aforementioned sources is proposed, which makes use of the streaming-data method integrated into the data processing engine indicated above and which is based on Machine Learning. This technique was used to analyze real-world data from smart-city IoT infrastructures as well as data from smaller-scale IoT deployments. In order to process the volume of data generated by the Internet of Things (IoT) in real time, several machines are required, despite the fact that this is true for the majority of data processing methods and methodologies applied in prior computer science application cases. This process began with the adoption of big data technologies such as Hadoop [14]. which were implemented in the early phases. A collection of actions that can be done consecutively to reduce the real number of data available to a modest and usable amount of information, and which were able to overcome the first challenges encountered with IoT systems were discovered. Although they have a viable solution for the most common IoT applications, they are unable to provide a good solution for scenarios when real data is not immediately available to the system that would process it a priori. What is known as "streaming" data is one of the most common examples of how data from Internet of Things installations would reach the processing system.

For managing the massive data streams generated by smart greenhouses using the Internet of Things, a number of research prototype solutions and commercially targeted systems have been developed. Because of the large number of requirements, including such publishing as well as subscribing to data streams, connecting with many technologies, and performing real-time analysis, it is necessary to create systems that expand many fundamental RDBMS systems. In this location, there was a flood detection and monitoring system available [15]. A secure, scalable platform enables actual control and storage for devices and items. IoT data processing solutions are also available from other firms, including such Amazon and Microsoft, who are primarily focused on delivering computing services, such as AWS IoT as well as Azure IoT Suite [16]. Device makers commonly are using a cloud-based data storage system or an analytics platform to store and analyze their data.

Messaging between clients and servers through a centralized service that also functions as a flood detector is at the heart of all of these methods. Updates on a list of available subjects maintained by a flood detection generator can be posted by client applications, and other applications can subscribe to receive updates on the list of available subjects. This design concept makes it possible to build asynchronous system designs that are not constrained by hardware and/or software constraints.. [17] are examples of large-scale smart metering deployments that used Machine



learning portals (for electricity and water) and other visualization tools to support the system and engage end-users to participate. 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 3. An advance home automation system schema developed for home using machine learning-technology [17].

As noted in [17], when it comes to 2030, we must place emphasis on the ability of all individuals to make educated and well-considered judgments. Furthermore, we can't manage what we can't measure, which is another fundamental reality. If we want a deeper knowledge of something like our daily energy use, we must keep track of the influence that our current behavior has on the situation, as well as the effect of potential behavioral alterations. Environmental education, which is a subset of science education in general [18], has played a crucial role in the development of Europe's cultural history. According to EU officials, environmental education is one of the most essential strategies for shifting human behavior away from harmful patterns and toward more environmentally friendly ones [19]. Integrating environmental education with game-based learning, as a result, allows students to take ownership of their own education by asking probing questions, sorting through possibilities, and creating hypotheses of their own.

## 6. METHODOLOGY

As of late we hear like never about web of things (Machine learning), fake power cognizance, and home computerization. These terms are comparative by the way they sound, yet they are different in alternate ways. This section will give the execution subtleties Machine learning based canny home robotization frameworks as a rule, and that incorporates, MQTT convention for hand lint the mechanization interaction with home server modules and for this reason wellspring of

programming language being utilized is python programming 3.8 on boa constrictor guide stage with both CPU and GPU use. The fundamental stream outline has been displayed in Figure 4 .

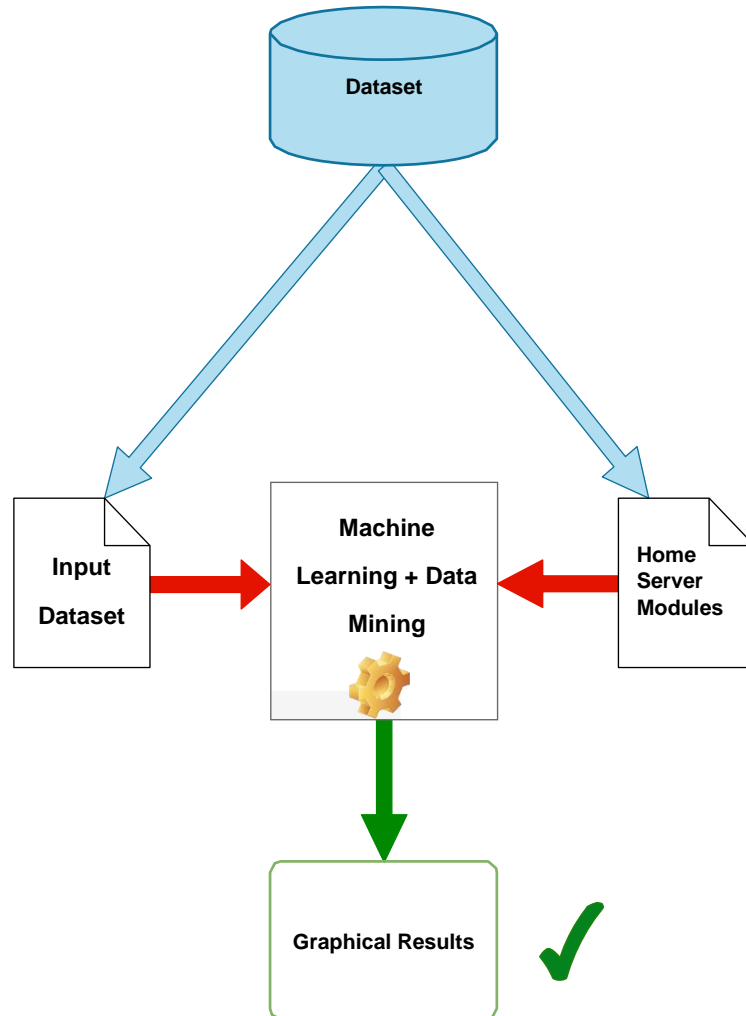


Figure 4. Flow diagram of approach being followed.

## 6.1. INTERNET OF THINGS ELEMENTS

Categories:

- a. Users: are the individuals who will interact with or live within the installation and are therefore referred to as such.
- b. Sensing Devices: These Internet of Things devices both gather and serve as a source of data. These are referred to as "sensing devices." It is feasible to transfer, consume, and save their data in a variety of ways.
- c. Actuator Devices: In the Internet of Things, actuator devices are devices that may be used to influence the physical world by sending commands to them. This category includes things like light switches and the thermostat settings on your HVAC system, to name a couple of examples.
- d. Gateway Devices: Gateway devices are Internet of Things devices that are solely responsible for the transport of information between the installation and the Internet (or vice versa). These devices play a critical role in installations where the other Internet of Things devices deployed are not Internet-enabled and hence unable to communicate with each other or the rest of the system.

e. Observed Phenomena: When we talk about "seen phenomena," we're referring to the physical phenomena that can be observed through the usage of digital sensing and actuator devices.

f. Units of The Measurements: Units of Measurements refer to the convention employed to quantify an observed occurrence.

Locations: Locations relate to the physical and logical groupings that can be applied to any of the elements presented above. This language can be used to identify not only the physical location of a building, school department, or classroom, but also an abstract assembly of persons scattered over several institutions. In a similar fashion there are linkages between the items indicated above that are defined as follows.

## 6.2. DATASET DESCRIPTION

For this exploration, REFIT: Electrical Load Measurements has been obtained from an open-source archive. The exploration worked on this dataset on python programming stage. The datasets documents reflect how shrewd home electrical framework act in space; hubs have sight of contacts with one another now and again for specific timeframes spans. Those Datasets reflect hubs conduct inside their power need, regardless of whether they are dynamic (the hub gets associated or disengaged because of its development), or latent (the hub is fixed, yet it gets associated or separated because of different hubs development), or a blend of both. The dataset remembers cleaned electrical utilization information for Watts for 20 families at total and machine level, timestamped and tested at 8 second stretches. The explanation for utilizing more modest dataset is that there is no requirement for large dataset assuming the objective is simply testing to contrast exhibitions rather than preparing with arrive at high precision. Besides, more modest dataset implies less an ideal opportunity to perform tests and thus additional time accessible to do significantly more investigations. The dataset could be downloaded from the link: <https://pureportal.strath.ac.uk/en/datasets/refit-electrical-load-measurements-cleaned>.

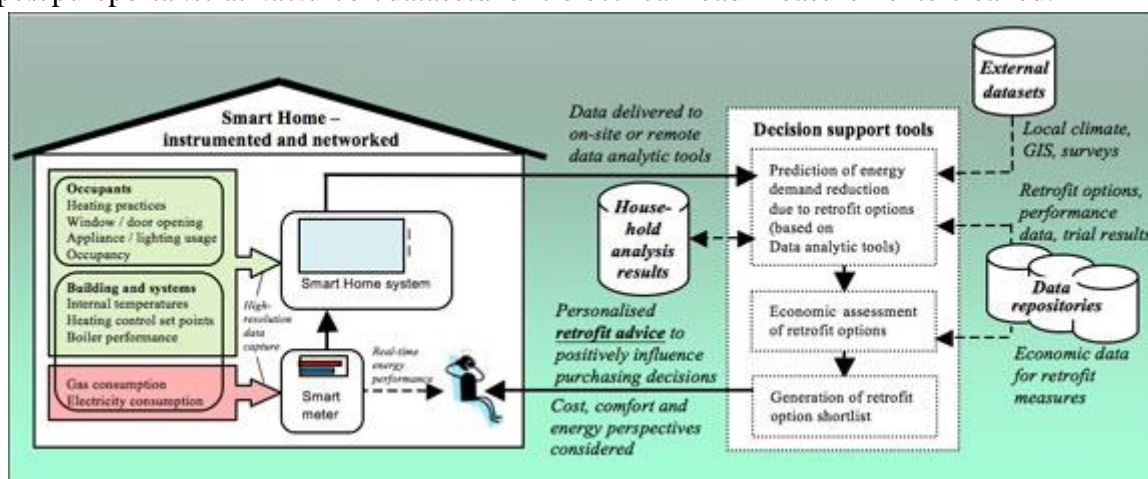


Figure 5. The dataset as a smart home instrumented and networked for analysis containing the data repositories

## 7. RESULTS

Our system is expected to allow the usage of a device with low memory and taking care of force limits. Our requirements are in this manner irrelevant and our system's unpredictability is on the home server side, that is the explanation a device should simply frequently consider its parts and properties. itself, fundamentally diminishing its hardware necessities. The base necessities are



presented and basically a device can communicate with the web over WiFi: that contraption ought to execute the TCP/IP show and have the choice to do HTTP interest/response due to the show. MQTT and home server movement. An additional a part that our system grants is the development of ports for partner more fundamental devices with no base essentials. The thinking is that the client connects with no less than one fundamental devices, for instance a fairly more amazing device. Arduino with WiFi receiving wire and this contraption goes probably as the least demanding section to our keen home structure. This is possible in light of the fact that in our structure we can without a very remarkable stretch add or dispose of novel contraptions and a port in the correspondence is direct to the whole system. This port ought to be enough good to inside manage the principal contraption information, for instance internal depiction and internal arranging used in the different correspondence show. This structure allows any client to be used on any stage: be it a wireless application, smartwatch application, site or whatever else. The principle need is that the application ought to have the choice to execute the HTTP show, as the client ought to have the choice to make RESTful API choices. Similarly, for devices, if the client has a JSON library, the normal code is more clear, yet not required. The clients of our system can be of two particular sorts: other Machine learning structures and for the present circumstance they don't give a UI (this is an ordinary M2M correspondence circumstance) or they are expected to give client association, and for the present circumstance, they should be arranged by usability best practices. In the execution of our system we didn't make any UI (out of degree), we just used the request line to test the REST API calls of the client, which This shows that the client least essentials.

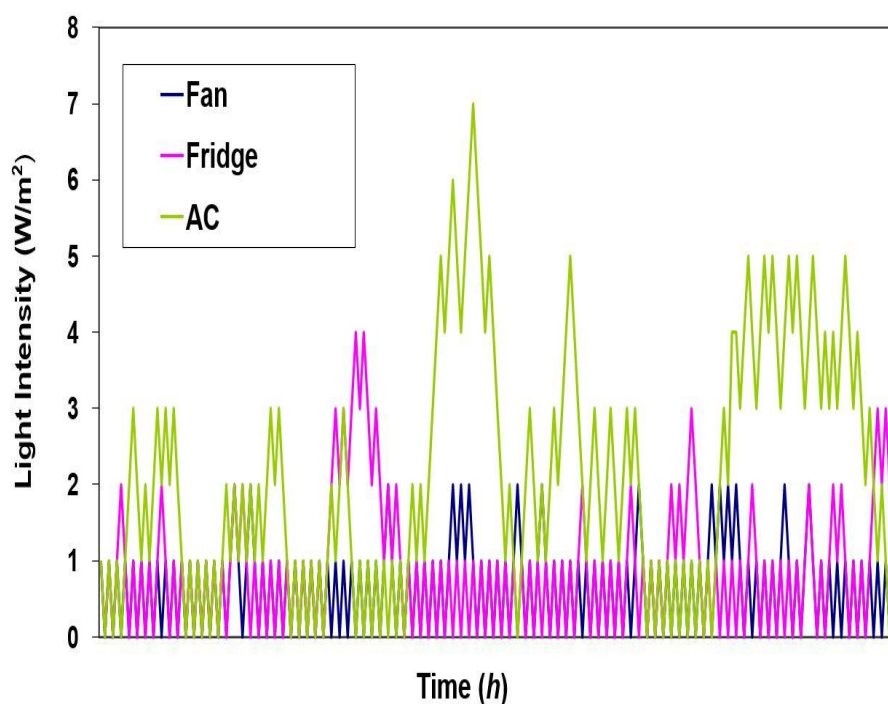


Figure 6: The light usage intensity of different appliances in home with respect to time.

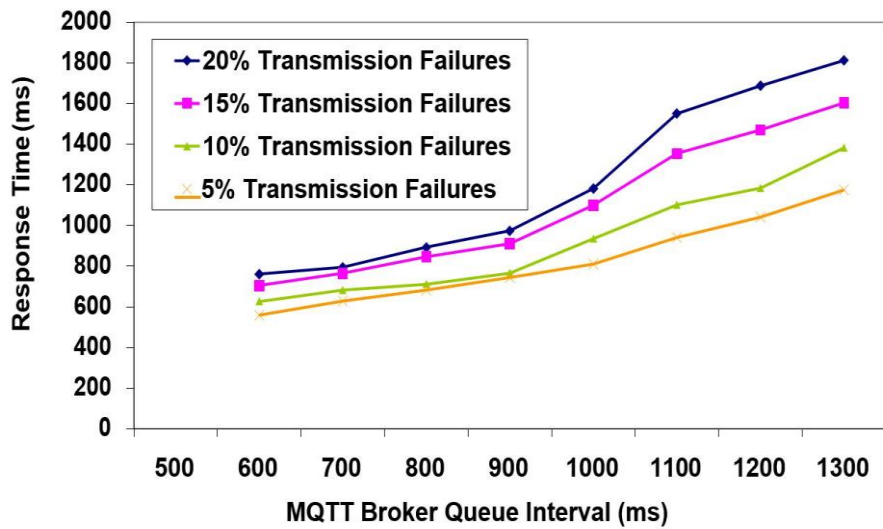


Figure 7. The transmission failure percentage for MQTT broker queue in property effecting the response time.

To address how the system capacities, we will explain a delineation of a structure made by one client, one home server and its Machine learning server, one MQTT middle person and one device with two sections, an Orange bar for the 5% transmission disillusionment and a Blue bar for the 20% transmission dissatisfaction. All of these parts have one property, mode, whose values can be 0 (OFF) or 1 (ON). This model blueprints how a contraption should plan their enlistment and how the MQTT vendor can change a sun controlled energy regard.

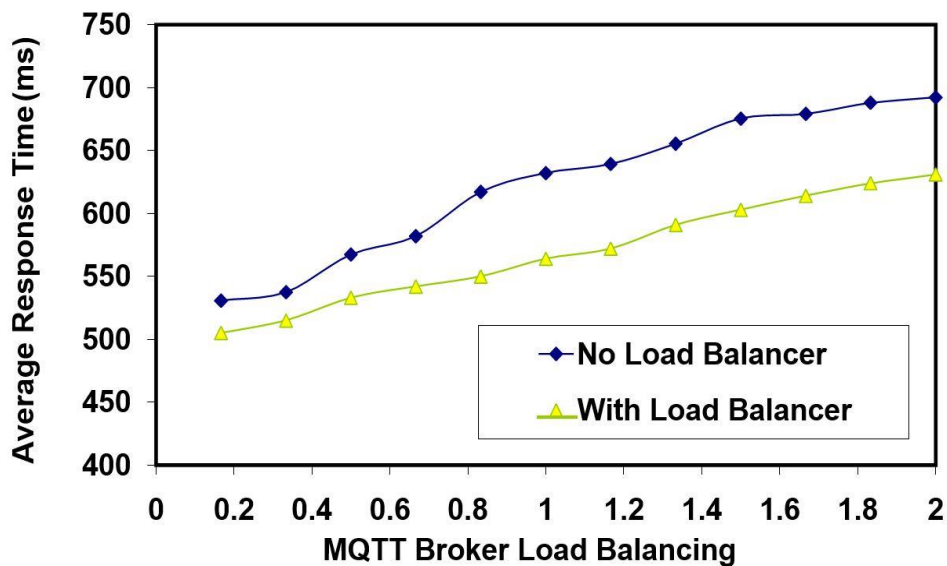


Figure 8. The load balancing of MQTT protocol reduces the average response time of electronics and appliances with respect to no load balancing.

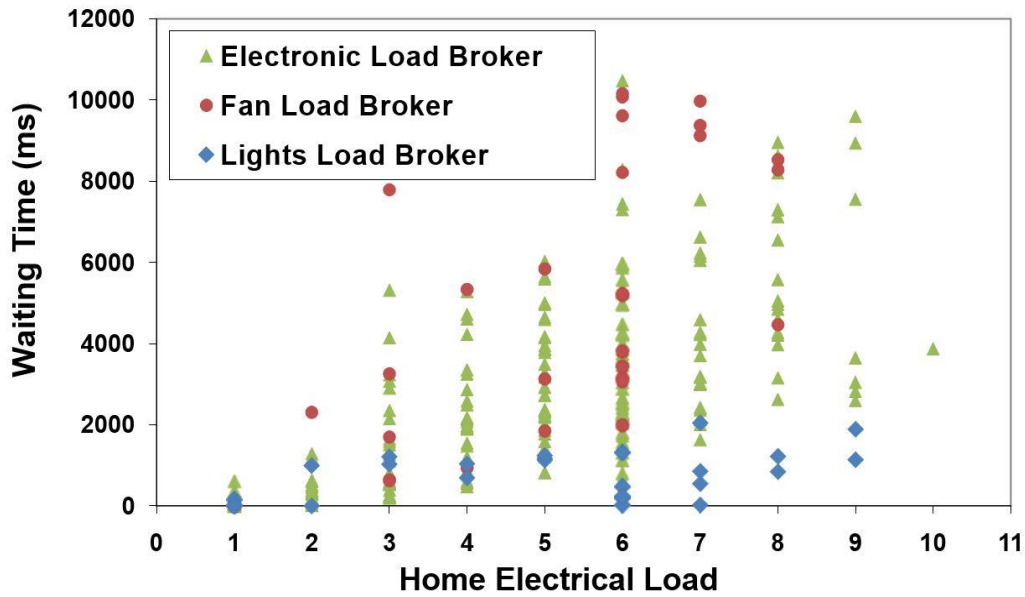


Figure 9 .This graph explains the electrical home load for fan, lights and electronic devices with the waiting time required for re-configuration through restful services.

To have a useful savvy home framework, it is vital that the home electrical burden arranges the framework as indicated by its necessities. To accomplish that, the MQTT intermediary ought to design, in any event, the accessible rooms and portions of home with sun oriented energy.

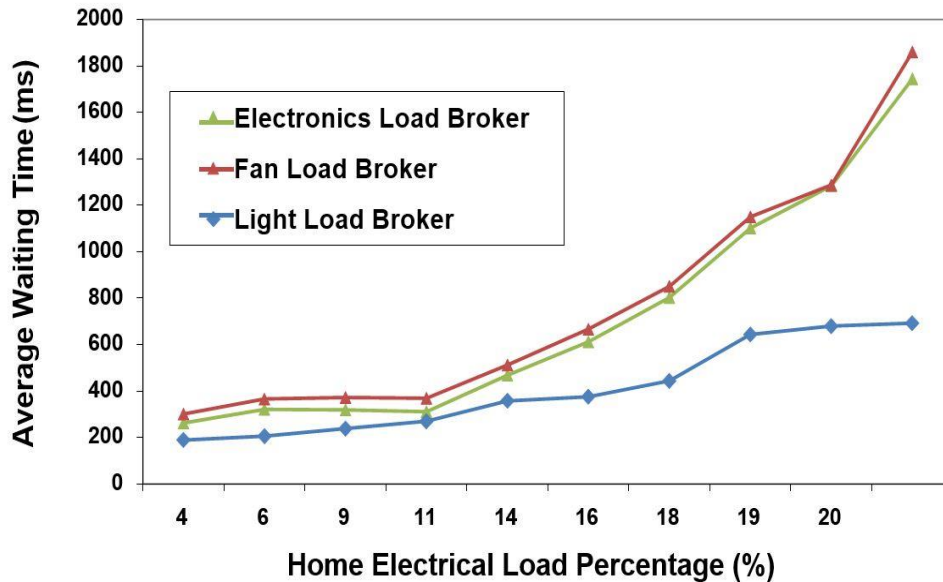


Figure 10. The percentage of home electrical load with respect to the average waiting time, the light load broker percentage has lowest effect.

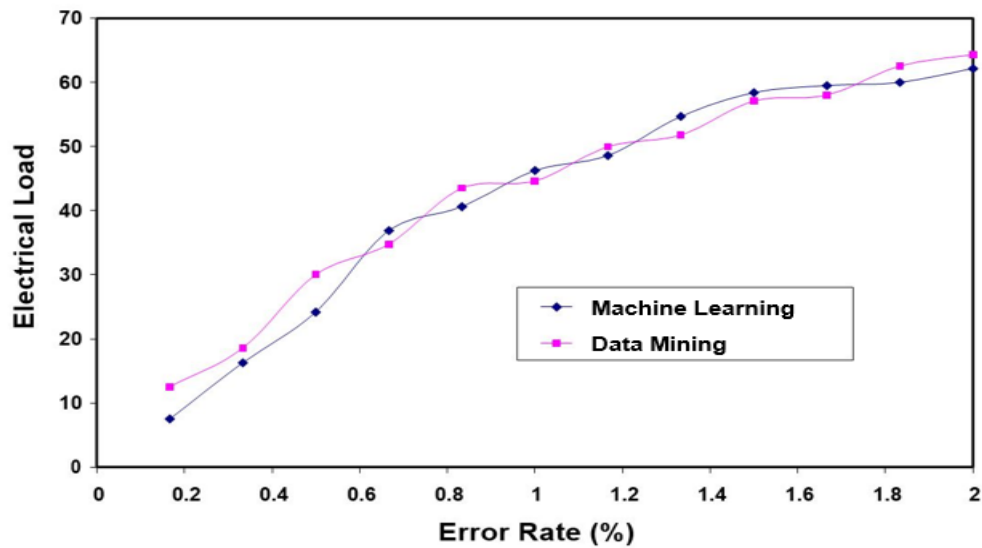


Figure 11. A comparison between the machine learning protocol and data mining protocol for handling the electrical load of home with respect to low error rate in percentage.

## 8. CONCLUSION

A machine learning-basically based brilliant homegrown construction with sun energy following is provided on this examination, it transformed into planned as a multi-facet structure with extraordinary parts, developed to be without trouble conveyed without requiring a convoluted framework. The secluded design guarantees a decoupling among the shrewd homegrown focus gadget and the devices connected, granting adaptability for the clients to substitute them and the brilliant homegrown gadget arrangements with regards to their requirements, with the gadget currently in activity. Indeed, even eleven however our evaluation transformed into done the use of handiest one instrument with MQTT convention connected with reproduction being done withinside the python programming, in an overall security appraisal, the gadget accomplishes the proposed objectives in general. It has a simple MQTT convention that helps a local area of various Machine learning devices connected and it has also fundamental abilities for use in genuine homegrown situations, for example the Machine learning consents component. The WiFi people group, on the grounds that the middle verbal trade age in our gadget, with the MQTT convention for verbal trade among devices transformed into an excellent decision, as it decreases the base necessities for the contraptions and grants to append devices with low computational sun energy sources to act in our gadget. Likewise, we took gain of a couple of abilities of the MQTT convention to guarantee insurance and interoperability in our gadget, which incorporate the QoS levels that we use to determine the messages significance, wherein with regards to the portrayed levels, the messages ought to have transport confirmation. The held message gadget, also reused from the MQTT convention, permitted new devices to get information roughly the elective contraptions' country right away when they participate in our gadget. The choice to offer contributions for machine learning contraptions in savvy homegrown to join from an external perspective, through the Internet, transformed into extremely fundamental as it isolates the external local area from the inward local area, developing the security and limiting the gets to the devices.

## REFERENCES

1. Cho, J.Y.; Kim, K.-B.; Jabbar, H.; Sin Woo, J.; Ahn, J.H.; Hwang, W.S.; Jeong, S.Y.; Cheong, H.; Yoo, H.H.; Sung, T.H. Design of optimized cantilever form of a piezoelectric energy harvesting system for a wireless remote switch. *Sens. Actuators A* 2018, 280, 340–349.
2. Mallick, S.; Bin Habib, A.-Z.S.; Ahmed, A.S.; Alam, S.S. Performance appraisal of Wireless Energy Harvesting in Machine learning. In *Proceedings of the 2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, Khulna, Bangladesh, 7–9 December 2017.
3. Amaro, J.P.; Cortesão, R.; Ferreira, F.J.T.E.; Landeck, J. Device and operation mechanism for non-beacon IEEE802.15.4/Zigbee nodes running on harvested energy. *Ad Hoc Networks* 2015, 26, 50–68.
4. Bianchi, V.; Ciampolini, P.; De Munari, I. RSSI-Based Indoor Localization and Identification for ZigBee Wireless Sensor Networks in Smart Homes. *IEEE Trans. Instrum. Meas.* 2018, 1–10.
5. Al-Sarawi, S.; Anbar, M.; Alieyan, K.; Alzubaidi, M. Internet of Things (Machine learning) communication protocols: Review. In *Proceedings of the 2017 8th International Conference on Information Technology (ICIT)*, Amman, Jordan, 17–18 May 2017. 38. QCA4020|Qualcomm. Available online: <https://www.qualcomm.com/products/qca4020>
6. Bassoli, M.; Bianchi, V.; De Munari, I.; Ciampolini, P. An Machine learning approach for an AAL Wi-Fi-based monitoring system. *IEEE Trans. Instrum. Meas.* 2017, 66, 3200–3209.
7. Moraru, S.; Perniu, L.; Kristaly, D.M.; Ungureanu, D.A.; Sandu, F.; Mo, A.A. *Proceedings of the 2017 25th Mediterranean Conference on Control and Automation (MED)*, Valletta, Malta, 3–6 July 2017.
8. Kristaly, D.M.; Moraru, S.-A.; Stefan Petre, V.; Parvan, C.A.; Ungureanu, D.E.; Mosoi, A.A. A Solution for Mobile Computing in a Cloud Environment for Ambient Assisted Living. In *Proceedings of the 2018 26th Mediterranean Conference on Control and Automation (MED)*, Zadar, Croatia, 19–22 June 2018.
9. Lago, P.; Roncancio, C.; Jiménez-Guarín, C. Learning and managing context enriched behavior patterns in smart homes. *Futur. Gener. Comput. Syst.* 2019, 91, 191–205.
10. Bassoli, M.; Bianchi, V.; De Munari, I.; Ciampolini, P. An unobtrusive Wi-Fi system for human monitoring. In *Proceedings of the IEEE International Conference on Consumer Electronics*, Berlin, Germany, 3–6 September 2017.
11. Boztas, A.; Riethoven, A.R.J.; Roeloffs, M. Smart TV forensics: Digital traces on televisions. *Digit. Investig.* 2015, 12, 72–80.
12. Chung, H.; Park, J.; Lee, S. Digital forensic approaches for Amazon Alexa ecosystem. *Digit. Investig.* 2017, 22, 15–25.
13. Orr, D.A.; Sanchez, L. Alexa, did you get that? Determining the evidentiary value of data stored by the Amazon R Echo. *Digit. Investig.* 2018, 24, 72–78.
14. Fensel, A.; Tomic, D.K.; Koller, A. Contributing to appliances' energy efficiency with Internet of Things, smart data and user engagement. *Future Gener. Comput. Syst.* 2017, 76, 329–338.
15. Goudbeek, A.; Choo, K.K.; Le-Khac, N.A. A Forensic Investigation Framework for Smart Home Environment. In *Proceedings of the 2018 17th IEEE International Conference on Trust*,



- Security and Privacy in Computing And Communications/12th IEEE International Conference on Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, USA, 1–3 August 2018; pp. 1446–1451.
16. Alharbi, R.; Allen, W.H. Collection and Analysis of Digital Forensic Data from Devices in the Internet of Things. In Proceedings of the 2019 SoutheastCon, Huntsville, AL, USA, 11–14 April 2019.
  17. Islam, M.J.; Mahin, M.; Khatun, A.; Debnath, B.C.; Kabir, S. Digital Forensic Investigation Framework for Internet of Things (Machine learning): A Comprehensive Approach. In Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 3–5 May 2019; pp. 1–6.
  18. Philomin, S.; Singh, A.; Ikuesan, A.; Venter, H. Digital Forensic Readiness Framework for Smart Homes. In Proceedings of the International Conference on Cyber Warfare and Security, Islamabad, Pakistan, 29 September–1 October 2020; pp. 627–638.
  19. MacDermott, A.; Baker, T.; Shi, Q. Machine learning forensics: Challenges for the ioa era. In Proceedings of the 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Paris, France, 26–28 February 2018; pp. 1–5.