





## Article

# Discovery of Resident Behavior Patterns Using Machine Learning Techniques and IoT Paradigm

Josimar Reyes-Campos <sup>1</sup>, Giner Alor-Hernández <sup>1,\*</sup>, Isaac Machorro-Cano <sup>1</sup>, José Oscar Olmedo-Aguirre <sup>2</sup>, José Luis Sánchez-Cervantes <sup>3</sup> and Lisbeth Rodríguez-Mazahua <sup>1</sup>

<sup>1</sup> Tecnológico Nacional de México/I. T. Orizaba, Av. Oriente 9,852, Col. Emiliano Zapata, Orizaba 94320, Veracruz, Mexico; josi.reyescampos@gmail.com (J.R.-C.); imachorro@gmail.com (I.M.-C.); lrodriguez@orizaba.tecnm.mx (L.R.-M.)

<sup>2</sup> Department of Electrical Engineering, CINVESTAV-IPN, Av. Instituto Politécnico Nacional 2,508, Col. San Pedro Zacatenco, Delegación Gustavo A. Madero, Mexico City 07360, Mexico; oolmedo@cinvestav.mx

<sup>3</sup> CONACYT-Tecnológico Nacional de México/I. T. Orizaba, Av. Oriente 9,852, Col. Emiliano Zapata, Orizaba 94320, Veracruz, Mexico; jlsanchez@conacyt.mx

\* Correspondence: galorh@orizaba.tecnm.mx; Tel./Fax: +52-272-725-7056

**Abstract:** In recent years, technological paradigms such as Internet of Things (IoT) and machine learning have become very important due to the benefit that their application represents in various areas of knowledge. It is interesting to note that implementing these two technologies promotes more and better automatic control systems that adjust to each user's particular preferences in the home automation area. This work presents Smart Home Control, an intelligent platform that offers fully customized automatic control schemes for a home's domestic devices by obtaining residents' behavior patterns and applying machine learning to the records of state changes of each device connected to the platform. The platform uses machine learning algorithm C4.5 and the Weka API to identify the behavior patterns necessary to build home devices' configuration rules. Besides, an experimental case study that validates the platform's effectiveness is presented, where behavior patterns of smart homes residents were identified according to the IoT devices usage history. The discovery of behavior patterns is essential to improve the automatic configuration schemes of personalization according to the residents' history of device use.

**Keywords:** behavior patterns; comfort; domestic; Internet of Things; machine learning



**Citation:** Reyes-Campos, J.; Alor-Hernández, G.; Machorro-Cano, I.; Olmedo-Aguirre, J.O.; Sánchez-Cervantes, J.L.; Rodríguez-Mazahua, L. Discovery of Resident Behavior Patterns Using Machine Learning Techniques and IoT Paradigm. *Mathematics* **2021**, *9*, 219. <https://doi.org/10.3390/math9030219>

Academic Editors: Matei Oliviu and Andreica Anca

Received: 24 December 2020

Accepted: 16 January 2021

Published: 22 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Nowadays, machine learning techniques have gained importance in various study areas due to their large number of pattern discovery applications. Under the changing conditions of the environment, individuals tend to develop behavioral patterns to better adapt to their environment. The analysis of the observed actions and their influences in the environment may lead to the automatic recognition of those meaningful behavioral patterns. Machine learning methods are appropriate when observable actions are available though there are no precise specifications for the system's desired behavior. There are currently home automation systems, such as Google Home or TESSY Cloud, that allow remote monitoring and control of home devices, while others have automatic configuration schemes, such as Google Home Routines. However, most of them do not consider user behavior analysis to discover patterns of use to establish fully personalized automatic configurations adjusted to each home's unique context. In those systems, automation occurs rigidly without considering changes or transformations in users' behavior as time passes. The automatic configuration is essential since, by counting on the users' behavior patterns, it will be possible to establish the appropriate arrangements adjusted to the inhabitants' preferences or needs at home. The configuration will lead to an increase in

the comfort of the lifestyle. In turn, it is possible to strengthen the new configurations or personalized domotic control rules with energy rules, which implies a balance between comfort and energy care. In the Encyclopedia of Computational Science, Reilly et al. [1] defined machine learning as the study and analysis of systems' desired behavior to develop and improve software systems without programming them directly.

Consequently, it is possible to take advantage of machine learning techniques to discover behavior patterns of people who live in a home by analyzing home domotic devices' records, aiming to automate efficiently the fully customized daily home setup functions. Having highly customizable automatic configuration schemes by analyzing user behavior allows, among other things, to improve residents' comfort schemes. However, the behavior analysis of users with their devices allows not only to improve comfort but also alleviate economic concerns from the point of view of energy-saving since it is possible to identify bad practices of daily energy consumption in residential or commercial environments. Energy-saving is a matter of high interest for society due to the great concern that exists to reverse the negative impact on the environment derived from the energy need caused by economic, recreational or other necessary activities [2].

IoT is an enabling technology that allows more and more devices to be accessed in real-time, as well as their energy and usage readings. IoT facilitates obtaining the necessary data to discover residents' behavior patterns of a home or commercial building. Krishna et al. [3] formally established that IoT consists of interconnected physical devices and software components intending to exchange information to provide a service to the end-user. In this way, IoT allows the strengthening of home automation areas, which comprises a set of methods and technologies designed to manage a house considering security, energy management, personal well-being and communication schemes [4].

This work presents Smart Home Control, an intelligent platform that offers fully customized automatic control schemes for domotic devices by obtaining residents' behavior patterns through the application of machine learning techniques to the records of state changes of each device connected to the platform. The machine learning algorithm C4.5 and the Weka API allow identifying the behavior patterns necessary to build the configuration rules for home devices. This work is an extension of that in [5], where the proposal of an energy-saving platform for smart homes considered the use of machine learning and IoT. Direct references for this paper are the works presented by Paredes et al. [6] and Machorro et al. [7], where they proposed home automation control through the implementation of IoT. Specifically, this work extends and strengthens the previous results by applying machine learning to discover behavior patterns that improve and customize the previously proposed automatic configuration schemes paying particular attention to improving comfort.

This paper's content is structured as follows. Section 2 discusses the papers related to machine learning, energy saving, IoT and smart homes. Subsequently, Section 3 presents the architecture of smart home Control and the case study applied to validate the platform's functionality. Section 4 presents the results and the discussion derived from them. Finally, Section 5 presents the conclusions and future work intentions.

## 2. Related Work

In recent years, IoT has been established as a technology that allows access to information from devices of daily use in a timely and efficient manner. In turn, the number of devices that now have an Internet connection has grown exorbitantly. Under this context, it is possible to monitor each IoT device's status, usage and energy consumption in a building, which provides excellent opportunities for improvement and strengthening of security, comfort and energy-saving schemes. The increment in the number and variety of IoT devices has revealed a lack of IoT platforms to manage and analyze such data. For this reason, an approach to limit the energy consumption that IoT devices cause when carrying out high data transmissions was presented by Azar et al. [8] to improve privacy and reduce network traffic while saving time. The approach proposed using edge computing, which allows downloading the workload from the cloud in a location closer to

the data source to be processed. To bring computing closer to where data are produced, Valerio et al. [9] performed data analysis on mobile nodes passing through IoT devices and explored the fog computing paradigm using a distributed machine learning framework (Hypothesis Transfer Learning). Saeid et al. [10] presented a taxonomy of machine learning algorithms that explains how different techniques are applied to data to extract higher-level information. They also evaluated machine learning methods that address the problems derived from the management of IoT data by considering smart cities as the case of primary use. Filho et al. [11] proposed a solution that combines computational intelligence and fog computing called STORM (Smart Solution for Decision Making in a Residential Environment). The STORM system, to improve the decision-making process, was able to detect and control the information generated by sensors installed in the residential scenario. Kasnesis et al. [12], within a collective intelligent environment and based on semantic Web technologies, proposed a platform that allows dynamic injections of automation rules. Fensel et al. [13] presented an IoT semantic platform over typical home appliance data called OpenFridge. Frontoni et al. [14] developed a framework to allow rapid development of complex systems by integrating new device classes into existing systems and controlling and centralizing information. Additionally, Silva et al. [15], to identify the appropriate solution for designing an IoT system, proposed a model and a framework. The results presented show that considering implementation time, cost, energy consumption, among other attributes, the methodology helped in the designing of an IoT system.

The reduction of energy consumption is a topic of considerable interest in the scientific community. Wen et al. [16] presented the ECIB algorithm that considers energy and cost variables to program IoT workflows with intensive batch processing in clouds. The main objective of ECIB is to improve energy efficiency and therefore reduce costs of operation. Another approach was proposed by Terroso-Saenz et al. [17] after introducing and testing the IoT Energy Platform (IoTEP), which aimed to provide, in terms of energy data management, an holistic solution for IoT energy data management. Pawar et al. [18], regarding power management, designed a smart system with the objective of replace, in a controlled manner based on consumer preference, a complete power outage in a region with a partial load outage. On the other hand, Zekić-Sušac et al. [19] used Rpart regression trees, random forest and variable reduction procedures to create predictive models of the specific energy consumption of public sector buildings in Croatia. The predictive models solve the problem of incorporating machine learning and the Big Data platform in the same intelligent system to manage the public sector's energy efficiency. Nabizadeh et al. [20] proposed an IoT-based Smartphone Energy Assistant (iSEA) framework that drives energy-conscious behaviors in commercial buildings. iSEA uses a deep-learning approach to identify individual occupants' energy use through their smartphone tracking and offers personalized feedback to impact their usage. García-Martín et al. [21] presented a review of the different approaches to estimating energy consumption on the one hand and machine learning applications on the other hand. Based on the data collected from an energy monitoring platform at a university in south China, Li et al. [22] proposed energy consumption patterns using data mining approaches. The coupling relationships between these components were revealed using multiple machine learning methods to propose precise energy-saving strategies for air conditioners' random use. Using data from temperature and humidity sensors, Raza et al. [23] developed machine learning models of heat flow of the environment and coming from the operation of heating, ventilation and air conditioning (HVAC) equipment. From these data, they obtained a low-cost and non-intrusive methodology for determining energy waste based on consumer behavior (CBB-EW) in HVAC operation and control. Chacon-Troya et al. [24] explained the design of an intelligent residence application for the control and monitoring of electricity quotas. To estimate the costs of the residence devices, they combined a Web application native technology. Aiming to reduce energy consumption, Saba et al. [25] contributed to the modeling and simulation of multi-agent systems for hybrid renewable energy system residences. In [26], Buono et al. presented The Non-Intrusive Load Monitoring (NILM)

System as a mobile application that shows historical and real-time information about energy consumption and sends message alerts whenever an energy overload is about to occur.

To strengthen an energy alert system that reduces costs to users, Weixian et al. [27] presented a home management self-learning system that classify and learn from smart homes data through machine learning. In [28], Elkhorchani and Grayaa described a system that reduces not only CO<sub>2</sub> emissions but also energy consumption rates. The home energy management system proposed by them employs an architecture that includes wireless communication, renewable energy principles and a power outage algorithm. Matsui [29] proposed a system that suggests ways to reduce electricity consumption through a module that also provides related information to be displayed according to the user's comfort preferences. Al-turjman et al. [30] reviewed and categorized existing energy monitoring approaches in the literature, investigating the impact and effectiveness of these monitoring systems under stress. Baker et al. [31] developed a multi-cloud IoT service composition algorithm called E2C2. The E2C2 algorithm seeks and integrates the fewest possible IoT services to achieve the user's requirements to create an energy-conscious composition system. Aiming to promote energy-saving, previous works regarding nature-inspired metaheuristic algorithms have revealed substantial contributions in the field of energy demand predictions [32–35].

Reinforcement of safety measures of IoT applications has also been an important research topic. Based on a deep migration learning model, Li et al. [36] proposed an intrusion detection algorithm (IDS) that helps identify network anomalies and takes necessary countermeasures to ensure IoT's safe and reliable operation applications. Alli et al. [37] proposed a secure calculation download scheme in the Fog-Cloud-IoT environment (SecOFF-FCIoT). Using machine learning strategies, they achieved an efficient and safe download in the Fog-IoT environment. Lanford and Perez [38], using video streaming, implemented a security system able to monitor the environment of a smart house. Malina et al. [39] presented a security framework that improves Internet of Things security and privacy services for the Message Queuing Transport Telemetry Protocol (MQTT).

In contrast, Mozaffari et al. [40] took into account a more human aspect of IoT devices. They considered the three stages of falling people (prediction, prevention and detection) to develop a diagnostic system for falls in smart buildings. At the same time, Ud Din et al. [41] studied different machine learning-based mechanisms on IoT used in healthcare, smart grids and vehicular communications to explain the role that machine learning techniques are expected to play in IoT networks development. In turn, Nižetić et al. [42] conducted a review of scientific contributions presented at the 4th International Conference on Smart and Sustainable Technologies held in Croatia, in 2019 (SpliTech 2019) to demonstrate the pros and cons of IoT technologies regarding an environmental point of view to promote the smart utilization of limited global resources.

As shown in this section, works related to this project's context were identified in the literature. However, of these works, only those in [11,12] are focused on improving the automatic configuration schemes. None considers the analysis or discovery of the residents' behavior patterns to improve the automatic control of domotic devices. Most papers expose issues related to energy-saving [8,9,13,16–35], security [36–39], data analysis [10], home system design [14,15] and healthcare [40–42]. After the reviewing these papers, a lack of smart configuration schemes that balance energy saving and comfort by analyzing user interaction with home automation devices was identified. Taking this into account, Smart Home Control is a solution capable of analyzing residents' historical interaction with their devices through a mobile application. From this analysis, behavior patterns are recognized to create automatic control rules that automate the home configurations and reduce energy consumption if necessary. Smart Home Control is a tool that improves and extends the functionality of a domotic system. The following section describes the architecture and functionality of Smart Home Control.

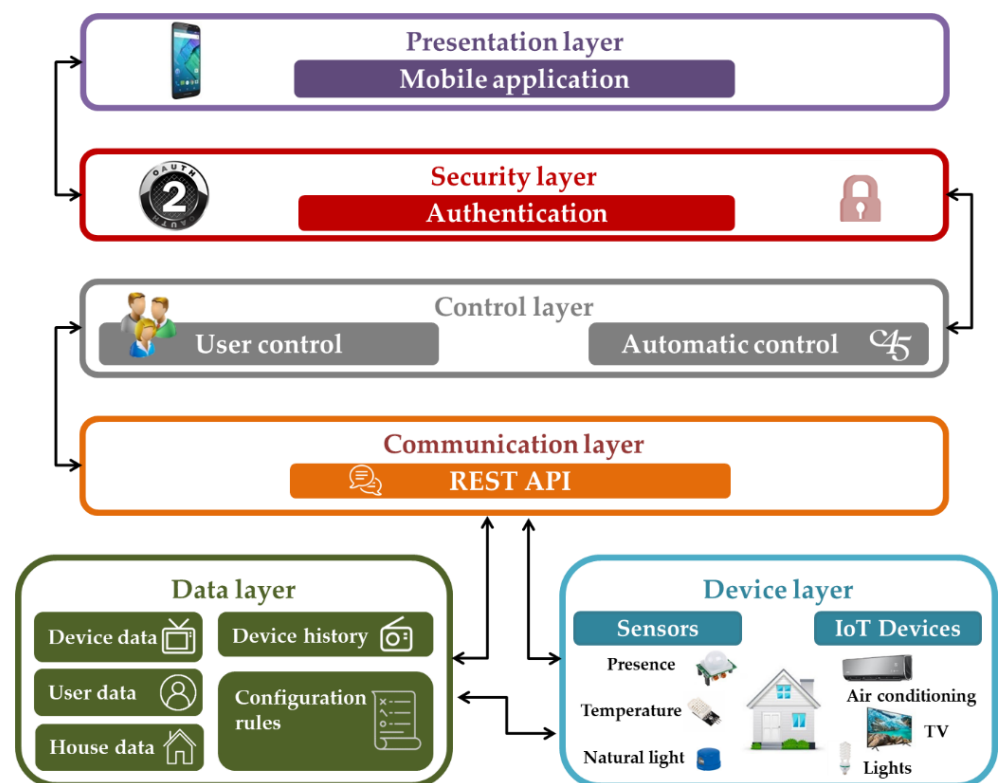
### 3. Materials and Methods

The need for a platform that analyzes the residents' behavior through machine learning arises from the lack of availability of a highly automated and personalized home appliance configuration schemes. Machine learning makes it possible to discover residents' behavior patterns, improving automated decision-making effectiveness as if the users' were controlling their devices manually. The "Smart Home Control" mobile application and its architecture were developed to meet this need.

#### 3.1. Smart Home Control's Architecture

The design of the Smart Home Control architecture includes a domotic system that contains the necessary technology to control and monitor IoT devices. The domotic system consists of devices with "IoT ready" technology installed in smart homes or buildings. In addition, as part of the home automation system, the harvesting of information from the surrounding environment regarding temperature and natural light detection has been considered. The Smart Home Control platform also includes a communication infrastructure that allows data to be obtained from sensors and sent to actuators or other IoT devices using web services.

The Smart Home Control architecture consists of six layers whose integration enables collecting and analyzing data from the IoT devices, providing control and communication for the home automation system. The layers that constitute the Smart Home Control are: (1) Presentation; (2) Security; (3) Control; (4) Communication; (5) Data; and (6) Devices. Figure 1 presents the Smart Home Control architecture, showing its layered design, the modules that comprise the layers and how they relate to each other.



**Figure 1.** Smart Home Control's architecture.

The functions for each element of the Smart Home Control architecture are described below.

- **Presentation layer** represents the interface that enables a user to interact with the Smart Home Control platform.

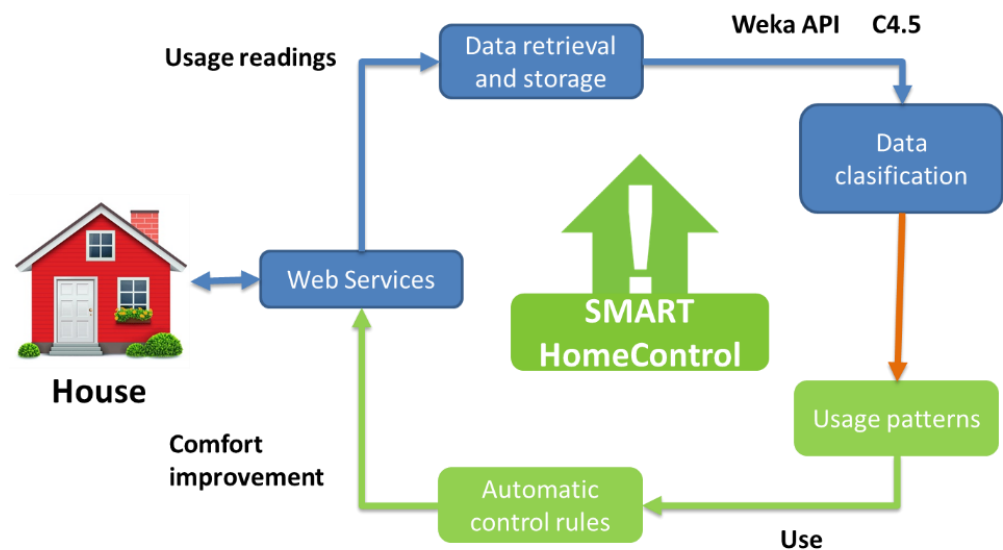


- **Mobile Application** is an Android-based interface through which a user gets access to the system. From this interface, the user can request to show real-time or historical data, select between automatic or manual control of the IoT devices and define their configurations.
- **Security layer** represents the set of technological elements whose activity is necessary to ensure secure access to the platform's functions.
  - **Authentication Module** is responsible for validating the user's access data with the information from the User Data module, through the implementation of the OAuth 2.0 authorization framework that enables a third-party application to obtain limited access to an HTTP service.
- **Control layer** represents the functionality of the platform. It contains the necessary methods to access each device's functions connected to the Smart Home Control platform.
  - **User control module:** A user has access to this module from the mobile application. From this module, the user can turn on, turn off, or change their devices' configuration settings. This module does not use prediction models.
  - **Automatic control module:** This module is in charge of processing the data from the devices' readings. The module analyzes the information from the devices of the house to discover the behavior patterns and conditions by recording some action or state changes. Said analysis is carried out using the C4.5 machine learning algorithm that can avoid data overclassification and subsequent overtraining. C4.5 can also handle both discrete and continuous incomplete data [43]. It is essential to mention that the behavior patterns detected allow generating decision trees that are translated into automatic configuration rules applicable to each device.
- **Communication layer** is the software infrastructure needed to establish communication between the various modules and elements of the Smart Home Control platform.
  - **REST API** contains methods that allow communication between the control layer, the device layer and the data layer. This layer collects data from sensor devices and transports the respective commands to control them.
- **Data layer** represents the information that Smart Home Control is focused on and the data of interest for the platform's various modules.
  - **Device data** corresponds to the stored data of each device connected to the platform. These data are mainly used to identify each device and its current status.
  - **Device history** is the historical information that includes each change in status registered by a specific device. This information is crucial because the record of all the devices' meaningful changes conveys the data necessary to analyze and discover the house inhabitants' behavior patterns. This cumulative information will effectively carry out the training of the automatic control module of the system.
  - **User data** corresponds to the stored data of the Smart Home Control users. These data are stored to personalize the user profile.
  - **House data** represents the stored data of the house where the Smart Home Control platform is installed. It includes information about the rooms to which home automation control devices are associated.
  - **Configuration rules** corresponds to the rules built by the automatic control module for each household device. The registered rules for a device may change as the automatic control module recognizes new device usage patterns.
- **Device layer** includes the communication technology necessary to control and monitor home automation devices. The intelligent system is implemented on this domotic system.

- **Sensors:** These are devices that detect and record changes in the environmental conditions (light, temperature, presence of people, etc.) for the home automation system.
- **IoT Devices:** These are devices connected to the domotic system through the use of IoT technology.

### Smart Home Control Workflow

A workflow for the Smart Home Control carries out the process of communicating with domotic devices, obtaining their data and analyzing them to build the automatic control rules of home automation devices. Figure 2 shows the Smart Home Control workflow, including the key concepts of the system's operation.



**Figure 2.** Smart Home Control's workflow.

The workflow is described as follows:

- The environment of a house should always provide access to domotic devices to obtain information of interest. It is advisable to consider an approach that includes IoT technology devices to facilitate communication and access to data corresponding to the history of use of the devices.
- The domotic devices, possibly having different communication interfaces, are accessed for reading and control purposes through web services. Each web service conforms to each device's specific communication interface, returning the information in the appropriate format so that the platform correctly stores the data of interest for later analysis.
- The data obtained from the devices are stored and subsequently analyzed by taking advantage of the Weka API that already has an implementation of the C4.5 algorithm, which aims to classify the collected data and thus allow the discovery of resident use patterns.
- After the usage patterns have been discovered, it is then possible to build decision trees that serve as a base for constructing custom comfort setting rules for each specific home's history. Consequently, each house where Smart Home Control is implemented will have appropriate rules fully adapted to the residents' conditions, allowing the automatic control to be dynamic by following an "always training–always evolving" approach.
- Finally, these new configuration rules will allow the platform to send the proper instructions to the devices through web services, thus improving the house's comfort conditions in terms of device control.

As can be seen, the central design idea of Smart Home Control is to follow an approach where the system remains in permanent learning. This design allows discovering the residents' usage behavior patterns, building control rules for home automation and readjusting these rules as usage patterns change.

### 3.2. Automatic Control Module: Pattern Discovery

The Smart Home Control Automatic Control module uses the Weka API 3.8 implementation of the machine learning algorithm C4.5 to analyze the information from IoT devices and discover the behavior patterns that allow generating the appropriate automatic control rules. The rules are applied to control devices with little or no human intervention, such as turning on the air conditioner when the thermometer reaches a temperature considered hot. Machine learning algorithms are used in many areas ranging from health services to financial frameworks. For this project, the C4.5 algorithm was selected for its performance and efficiency in building classification models that face prediction problems. There are previous works where the performance and effectiveness of various algorithms have been evaluated. Among them, the authors of [44] verified that the results of experimenting with the implementation of C4.5 lead to better levels of interpretability and precision of data. It also generates a shorter response time in the categorization processes compared to other algorithms such as Random Tree or fuzzy C-means. In [45], the authors confirmed the efficacy of C4.5 since after the tests were performed, finding that the algorithm was capable of making predictions with better performance metric results than other algorithms such as CART or even Random Forest. Data related to actions performed using Smart Home Control must be collected, stored and analyzed daily to discover usage patterns and generate rules. The C4.5 predictive algorithm uses the data to perform configuration automatically once the usage history of each device is available. Sensor readings, which can be numerical or nominal, are examined for behavioral patterns expressed in the form of decision trees that will be used to classify new readings [29] and control devices autonomously. Notice that the user's continuous use of Smart Home Control will allow obtaining better and more usage information, enabling the automatic configuration module to be trained with greater precision and producing more precise automatic configurations.

As mentioned above, residents can control their devices using the mobile application (created for the Android platform) of Smart Home Control. Constant interaction with the application allows Smart Home Control to obtain data that are used for the analysis and prediction of automatic settings for each device. Figure 3 presents some of the application manual configuration interfaces for a room and a single device.

Figure 4 shows the interfaces from which it is possible to set the whole house in intelligent control mode.

This automatic operation mode works according to the behavior patterns detected in the house. An example is when the lights or the air conditioner in the room should be turned on or off by considering temperature and human presence, among other settings.

Through the use of C4.5, the automatic configuration module analyzes the historical use of home automation devices to detect usage patterns that serve to identify the conditions of use of each device. The detection of these conditions allows the construction of classification rules for new values, which is useful in predicting future domotic configurations.



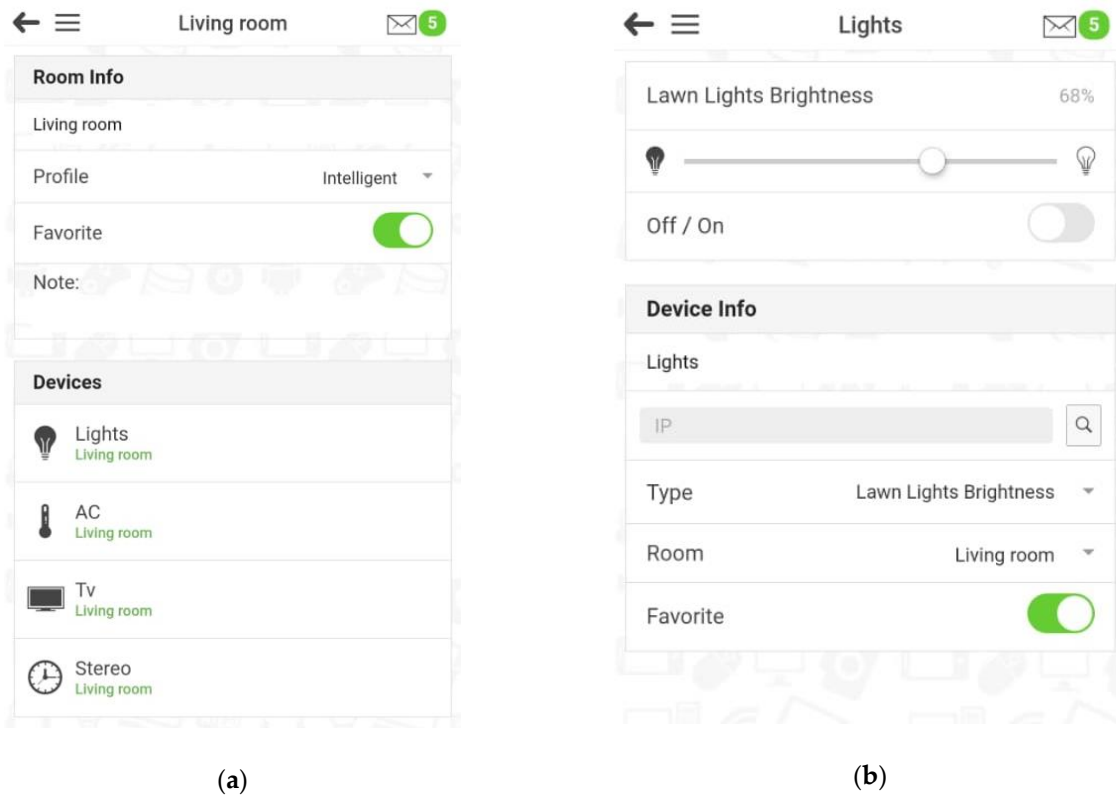


Figure 3. Manual configuration interfaces of the application: (a) room settings and its devices; and (b) device details and settings.

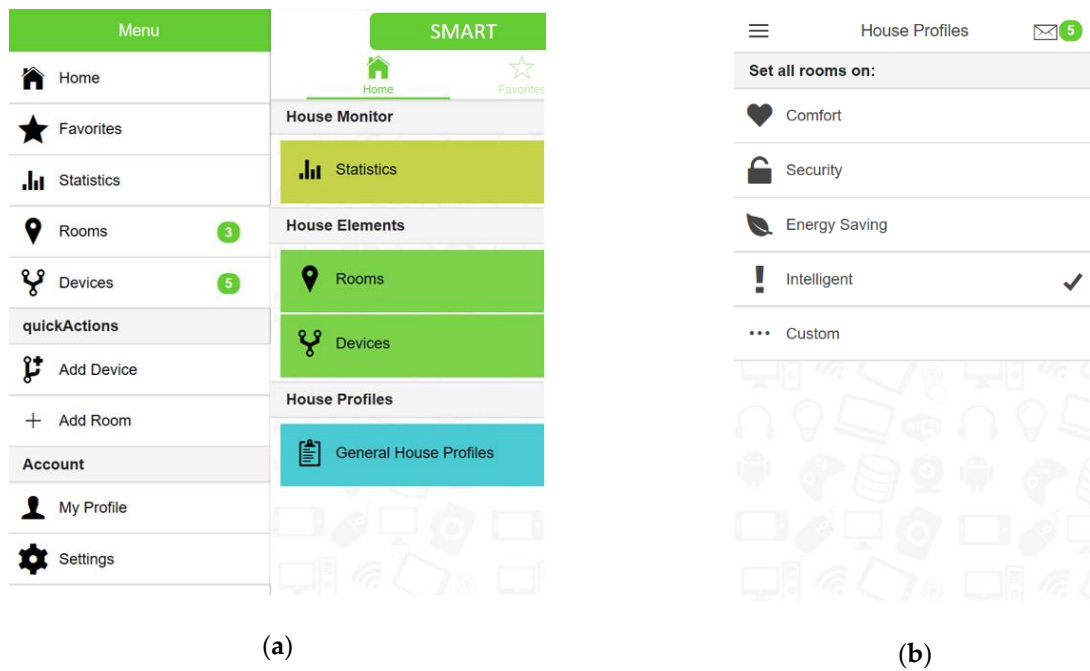
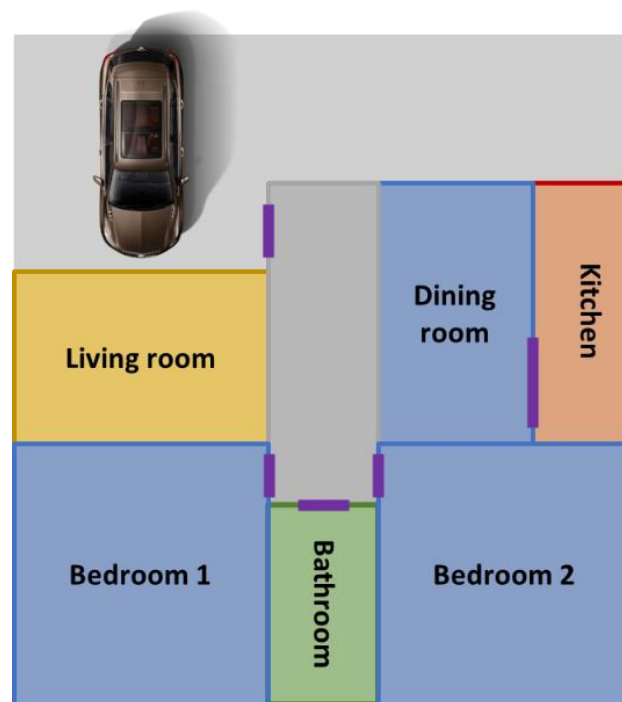


Figure 4. Intelligent mode interfaces: (a) main menu from which the configuration house profile options can be accessed; and (b) house profile selector.

### 3.3. Experimental Case Study

This section presents a case study results to validate that it is possible to discover patterns of home automation devices through Smart Home Control. This case study contemplates the data collected from the devices of the home automation system of a group of 10 houses located in Orizaba, Veracruz, Mexico. These houses were monitored during a period of 10 months (from March to December 2019). The training models of the system were defined from the information collected from the devices of each house. The construction of these training models is represented by a series of nested conditions in the decision tree format for each home automation device connected to Smart Home Control. Each device's decision tree is translated as a rule that allows its respective automatic on and off, which considers the house's general state. Figure 5 shows the house's distribution studied, which has a living room, dining room, kitchen, two bedrooms and a bathroom.



**Figure 5.** Visual representation rooms of the house.

In Figure 6, the scenario of the experimental case study is visually represented. It should be noted that the information of the house is accessed in real-time through sensors and devices of daily use with IoT technology-enabled. Smart Home Control uses web services to access devices and sensors to obtain input data that can be used to train the Automatic Control module. Once trained, the module is capable of controlling the devices as if the user did. Through a mobile application, the resident interacts with Smart Home Control with a user-friendly interface where notifications and information of interest from the devices connected to the system are displayed. The user is also able to control their devices manually through their mobile application.

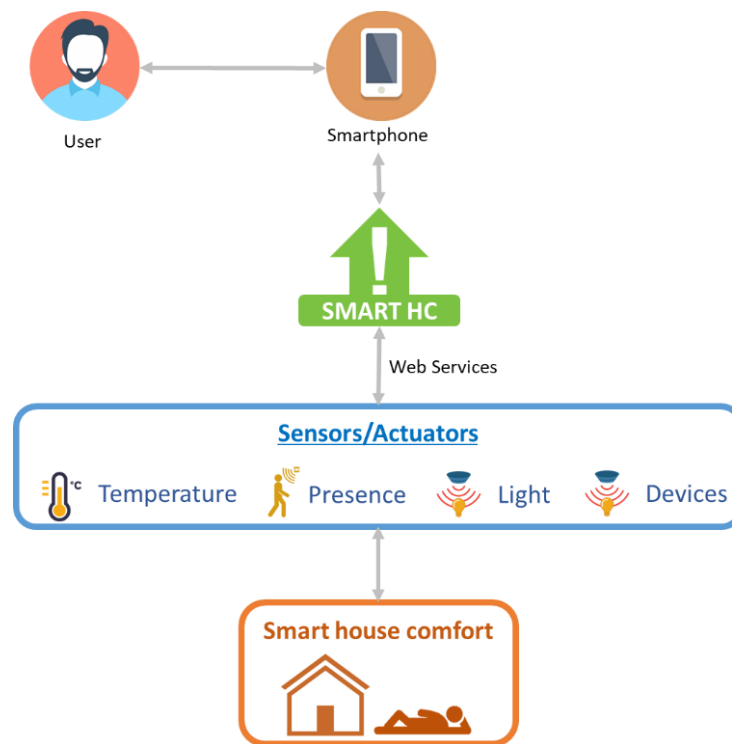


Figure 6. Visual representation of the scenario.

Three residents inhabited the house, and domotic devices were installed in each room. Additionally, sensors that allowed the recording of environmental conditions were installed to record possible variables to consider when taking any action on a specific device. For this reason, the on and off records of each of the devices were evaluated. The measurements recorded, e.g., human presence, natural light and temperature sensors, were considered. The registration of information from the sensors in each room is used to determine factors that could affect residents’ behavior patterns. These factors vary among the houses if variables such as geographic location, socioeconomic stratum, etc. are considered. For example, the case study was conducted in Veracruz’s central state area with a hot-humid climate. This climate encourages the frequent use of air conditioning devices, unlike other regions such as Mexico’s central area, where its cooler temperatures favor less use of these devices. Table 1 presents the devices, sensors and rooms taken into account for home automation registration and automatic control.

Table 1. Device distribution of the house.

Room	Dimensions of Room	Devices
Living room	3.00 m × 2.20 m	<b>Devices:</b> television, stereo, three lights, air conditioning <b>Sensors:</b> natural light, temperature, presence
Dining room	1.60 m × 3.00 m	<b>Devices:</b> three lights <b>Sensors:</b> presence, natural light
Kitchen	1.40 m × 3.00 m	<b>Devices:</b> three lights <b>Sensors:</b> presence, natural light
Bathroom	1.20 m × 2.40 m	<b>Devices:</b> two lights, odor extractor <b>Sensors:</b> presence, natural light
Bedroom 1	3.00 m × 2.80 m	<b>Devices:</b> television, two lights, two lamps, air conditioning <b>Sensors:</b> natural light, presence, temperature
Bedroom 2	3.00 m × 2.80 m	<b>Devices:</b> television, two lights, two lamps <b>Sensors:</b> natural light, presence, temperature

From March to December 2019, data collection and analysis were carried out, analyzing all household device configuration records including the readings of the sensors installed in each room. Table 2 presents part of the changes in the state of the devices and sensors installed in the dining room as an example of the information taken into account as training data.

**Table 2.** Some sensor data collected during the Experimental Case Study.

Day	Date	TV	Lights	AC	Temperature
Monday	07/09/2019	On	Off	Off	20.0 °C
Monday	07/09/2019	Off	Off	On	32.0 °C
Monday	07/09/2019	Off	Off	On	27.0 °C
Monday	07/09/2019	Off	On	Off	25.0 °C
Monday	07/09/2019	On	On	Off	19.0 °C
Monday	07/09/2019	Off	Off	On	25.0 °C
Tuesday	08/09/2019	On	On	On	23.0 °C
Tuesday	08/09/2019	Off	Off	On	23.0 °C
Tuesday	08/09/2019	On	Off	On	21.0 °C
Tuesday	08/09/2019	Off	Off	On	22.0 °C
Tuesday	08/09/2019	On	Off	On	27.0 °C
Tuesday	08/09/2019	On	On	On	21.0 °C
Tuesday	08/09/2019	Off	On	Off	21.0 °C
Tuesday	08/09/2019	Off	Off	Off	21.0 °C
Wednesday	09/09/2019	Off	Off	On	23.0 °C
Wednesday	09/09/2019	Off	Off	On	26.0 °C
Wednesday	09/09/2019	Off	Off	On	16.0 °C
Wednesday	09/09/2019	Off	Off	On	22.0 °C
Wednesday	09/09/2019	On	On	On	26.0 °C
Wednesday	09/09/2019	Off	On	On	21.0 °C
Wednesday	09/09/2019	Off	Off	On	22.0 °C
Wednesday	09/09/2019	On	Off	On	21.0 °C
Thursday	10/09/2019	On	Off	On	28.0 °C
Thursday	10/09/2019	On	Off	On	24.0 °C
Thursday	10/09/2019	On	Off	Off	24.0 °C
Thursday	10/09/2019	Off	Off	On	21.0 °C
Thursday	10/09/2019	Off	Off	On	21.0 °C
Thursday	10/09/2019	Off	Off	On	30.0 °C
Thursday	10/09/2019	On	On	On	23.0 °C
Thursday	10/09/2019	Off	On	Off	19.0 °C
Thursday	10/09/2019	Off	Off	Off	22.0 °C
Friday	11/09/2019	Off	Off	Off	20.0 °C
Friday	11/09/2019	On	Off	On	28.0 °C
Friday	11/09/2019	Off	Off	On	18.0 °C
Friday	11/09/2019	Off	Off	On	16.0 °C
Friday	11/09/2019	On	Off	On	22.0 °C
Friday	11/09/2019	Off	Off	On	20.0 °C
Friday	11/09/2019	Off	On	On	17.0 °C
Friday	11/09/2019	Off	Off	On	20.0 °C
Friday	11/09/2019	Off	Off	Off	22.0 °C

The following methodology was used to develop the experimental case study:

- Users were asked to manually control their domotic devices via the mobile app for 10 months, from March to December 2019.
- During this period, data were collected on changes in the state of the devices and sensors in the house that served as input values to train the automatic configuration module.
- The records of the state of all devices connected to Smart Home Control obtained throughout each day of the week were used to get the home automation usage patterns. The states of each device determined the global state of the house.

- Once the dataset was obtained, an analysis of each device's usage behavior was performed, taking into account the rest of the devices and sensors in the home.
- Each device was classified with possible "On" and "Off" values. Then, by using C4.5, decision trees were obtained whose conditions are based on the behavior observed during the training period. The rules resulting from the construction of the decision trees are dynamic and vary according to the particular behavior observed during the training.

The resulting control rules are the best at adjusting to the behavior of the house. The control rules allow automating the smart home to improve the comfort conditions of the users. Table 3 shows the attributes considered for the C4.5 analysis and pattern discovery.

**Table 3.** Attributes considered for the C4.5 analysis.

Attribute	Type	Device	Room	Description	Possible Values
TV @Livingroom	Nominal	Television	Living room	On/Off device log	["On", "Off"]
Stereo @ Livingroom	Nominal	Stereo	Living room	On/Off device log	["On", "Off"]
Lights @Livingroom	Nominal	Lights	Living room	On/Off device log	["On", "Off"]
NaturalLight @Livingroom	Nominal	Natural light sensor	Living room	Log of natural light in the room	["Yes", "No"]
AC @ Livingroom	Nominal	Air conditioner	Living room	On/Off device log	["On", "Off"]
Temp @Livingroom	Numeric	Temperature sensor	Living room	Temperature log (°C)	[Numeric value]
Presence @Livingroom	Nominal	Presence sensor	Living room	Log of human presence in the room	["Yes", "No"]
Presence @Diningroom	Nominal	Presence sensor	Dining room	Log of human presence in the room	["Yes", "No"]
Lights @Diningroom	Nominal	Lights	Dining room	On/Off device log	["On", "Off"]
NaturalLight@Diningroom	Nominal	Natural light sensor	Dining room	Log of natural light in the room	["Yes", "No"]
Presence @Kitchen	Nominal	Presence sensor	Kitchen	Log of human presence in the room	["Yes", "No"]
Lights @Kitchen	Nominal	Lights	Kitchen	On/Off device log	["On", "Off"]
NaturalLight @ Kitchen	Nominal	Natural light sensor	Kitchen	Log of natural light in the room	["Yes", "No"]
Presence @Bathroom	Nominal	Presence sensor	Bathroom	Log of human presence in the room	["Yes", "No"]
Lights @ Bathroom	Nominal	Lights	Bathroom	On/Off device log	["On", "Off"]
NaturalLight @ Bathroom	Nominal	Natural light sensor	Bathroom	Log of natural light in the room	["Yes", "No"]
Extractor @ Bathroom	Nominal	Odor extractor	Bathroom	On/Off device log	["On", "Off"]
TV @B1	Nominal	Television	Bedroom 1	On/Off device log	["On", "Off"]
Lights @B1	Nominal	Lights	Bedroom 1	On/Off device log	["On", "Off"]
Lamps @B1	Nominal	Lamps	Bedroom 1	On/Off device log	["On", "Off"]
NaturalLight @B1	Nominal	Natural light sensor	Bedroom 1	Log of natural light in the room	["Yes", "No"]
AC @B1	Nominal	Air conditioner	Bedroom 1	On/Off device log	["On", "Off"]
Temp @B1	Numeric	Temperature sensor	Bedroom 1	Temperature log (°C)	[Numeric value]
Presence @B1	Nominal	Presence sensor	Bedroom 1	Log of human presence in the room	["Yes", "No"]
TV @B2	Nominal	Television	Bedroom 2	On/Off device log	["On", "Off"]
Lights @B2	Nominal	Lights	Bedroom 2	On/Off device log	["On", "Off"]
Lamps @B2	Nominal	Lamps	Bedroom 2	On/Off device log	["On", "Off"]
NaturalLight @B2	Nominal	Natural light sensor	Bedroom 2	Log of natural light in the room	["Yes", "No"]
Presence @B2	Nominal	Presence sensor	Bedroom 2	Log of human presence in the room	["Yes", "No"]



As mentioned, to analyze each device's usage pattern in the system, it is necessary to consider the rest of the devices and sensors states of the house as variables. Each device analyzed is considered as a class attribute, which will be the classification target of C4.5. However, it is possible to analyze the numerical and nominal information collected from home automation devices to perform, for each class attribute, the selection of key attributes that allow the construction of decision trees with greater precision through the use of a correlation-based feature subset evaluator and the best first search method. Table 4 shows the selection of attributes considered for each class attribute of the experimental case study.

**Table 4.** Selected attributes for each class attribute.

<b>Class Attribute</b>	<b>Selected Attributes</b>
<b>TV @Livingroom</b>	Stereo @Livingroom Presence @Livingroom AC @Bedroom1 Presence @Bedroom2
<b>Stereo @Livingroom</b>	TV @Livingroom Lights @Livingroom TV @Bedroom1
<b>Lights @Livingroom</b>	Stereo @Livingroom Presence @Livingroom NaturalLight @Diningroom Lights @Kitchen Lights @Bathroom Lights @Bedroom2
<b>AC @Livingroom</b>	Temperature @Livingroom Presence @Livingroom Presence @Bedroom1
<b>Lights @Diningroom</b>	Stereo @Livingroom Presence @Diningroom NaturalLight @Diningroom Lights @Kitchen
<b>Lights @Kitchen</b>	Lights @Livingroom Lights @Diningroom Presence @Kitchen TV @Bedroom2
<b>Lights @Bathroom</b>	Lights @Livingroom Presence @Bathroom Extractor @Bathroom Lights @Bedroom1
<b>Extractor @Bathroom</b>	Presence @Bathroom Lights @Bathroom
<b>TV @Bedroom1</b>	Lights @Bedroom1 Lamps @Bedroom1 Presence @Bedroom1 Presence @Bedroom2
<b>Lights @Bedroom1</b>	NaturalLight @Diningroom Lights @Bathroom TV @Bedroom1 Lamps @Bedroom1 Presence @Bedroom1

Table 4. Cont.

Class Attribute	Selected Attributes
Lamps @Bedroom1	Temperature @Livingroom TV @Bedroom1 Lights @Bedroom1 NaturalLight @Bedroom1 Presence @Bedroom1 Lamps @Bedroom2
AC @Bedroom1	AC @Livingroom Lights @Diningroom NaturalLight @Diningroom NaturalLight @Kitchen Lights @Bathroom TV @Bedroom1 Lights @Bedroom1
TV @Bedroom2	Lights @Bedroom2 Presence @Bedroom2
Lights @Bedroom2	Lights @Livingroom NaturalLight @Diningroom TV @Bedroom1 TV @Bedroom2
Lamps @Bedroom2	Temperature @Livingroom NaturalLight @Bedroom1 Lamps @Bedroom1 Presence @Bedroom2

A set of metrics was selected to evaluate the C4.5 performance. These metrics are described as follows.

- **Precision** is, in the data classification field, the fraction of items correctly classified as positive out of all items that the algorithm classified as positive. The calculation considers true positive classification instances (tp) between the sum of true positive and false positive (fp) instances as
  - $precision = \frac{tp}{tp+fp}$ .
- **Recall** is the fraction of elements correctly classified as positive out of all the positive elements, and it is calculated considering true positive classification instances between the sum of true positive and false negative (fn) instances as
  - $recall = \frac{tp}{tp+fn}$
- **F-measure** provides a single score that balances both the concerns of precision and recall in one number, and it is calculated based on the precision and recall as
  - $Fmeasure = \frac{2 * precision * recall}{precision+recall}$
- **Matthews Correlation Coefficient (MCC)** measures the quality of binary (two-class) classifications taking into account true and false positives and negatives and returning values between  $-1$  and  $+1$ , where  $+1$  represents a perfect prediction and  $-1$  indicates a total discrepancy between observation and prediction. The MCC can be calculated using the formula
  - $MCC = \frac{tp*tn-fp*fn}{\sqrt{(tp+fp)(tp+fn)(tn+fp)(tn+fn)}}$
- **Receiver Operating Characteristics Area (ROC Area)** represents a classification model's performance at all classification thresholds. The PRC is a graph with: the x-axis showing False positive rate (FPR) calculated as
  - $FPR = \frac{fp}{fp+tn}$

and the y-axis showing True positive rate (*TPR*) calculated as

$$\blacksquare \quad TPR = \frac{tp}{tp+fn}$$

- **Precision–Recall Curve Area (PRC Area)** shows the relationship between precision and recall for every possible cut-off. The *PRC* is a graph with the x-axis showing recall and the y-axis showing precision

Table 5 presents the *precision*, *recall*, *F-Measure*, *MCC*, *ROC Area* and *PRC Area* metrics resulting from the device class analysis using C4.5.

**Table 5.** C4.5 performance metrics.

Class Attribute	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
TV @ Livingroom	0.742	0.747	0.743	0.44	0.785	0.78
Stereo Livingroom	0.791	0.788	0.79	0.396	0.82	0.837
Lights @Livingroom	0.833	0.848	0.838	0.405	0.604	0.782
AC @Livingroom	0.947	0.949	0.946	0.744	0.722	0.883
Lights @Diningroom	0.876	0.879	0.877	0.639	0.844	0.863
Lights @Kitchen	0.819	0.808	0.813	0.456	0.787	0.826
Lights @Bathroom	0.887	0.889	0.888	0.674	0.878	0.893
Extractor @Bathroom	0.99	0.99	0.99	0.978	0.986	0.977
TV @Bedroom1	0.806	0.818	0.81	0.394	0.674	0.77
Lights @Bedroom1	0.847	0.828	0.835	0.502	0.706	0.794
Lamps @Bedroom1	0.818	0.818	0.818	0.329	0.616	0.767
AC @Bedroom1	0.879	0.909	0.89	0.164	0.65	0.88
TV @Bedroom2	0.871	0.869	0.859	0.652	0.791	0.833
Lights @Bedroom2	0.911	0.909	0.91	0.733	0.933	0.921
Lamps@Bedroom2	0.835	0.838	0.837	0.594	0.789	0.789

Figure 7 shows an example of a classification tree that resulted from the data analysis of each of the house devices. The design of the trees was based on those presented by Pintelas et al. [46] who developed a semi-supervised methodology grey-box model to explain and understand how the predictive models works achieving a balance between the black box and white box paradigms. The tree shows the discovery of the usage patterns related to turning on and off the lights in the living room using algorithm C4.5. When considering the totality of data recorded by the rest of the devices and sensors, the resulting tree reveals a set of conditions applicable to the whole house.

Figure 8 shows a second example of a classification tree resulting from the analysis using algorithm C4.5 to set the rule for turning the living room air conditioner on and off.

Once the classification models have been obtained through the use of C4.5, it is possible for Smart Home Control to automatically turn on or off the devices previously described in this document. As the decision trees are built from the data recorded by the residents' spontaneous usage of home devices, the system can perform the actions as close as possible to the way the house inhabitants would do, representing an improvement in comfort in a smart home.

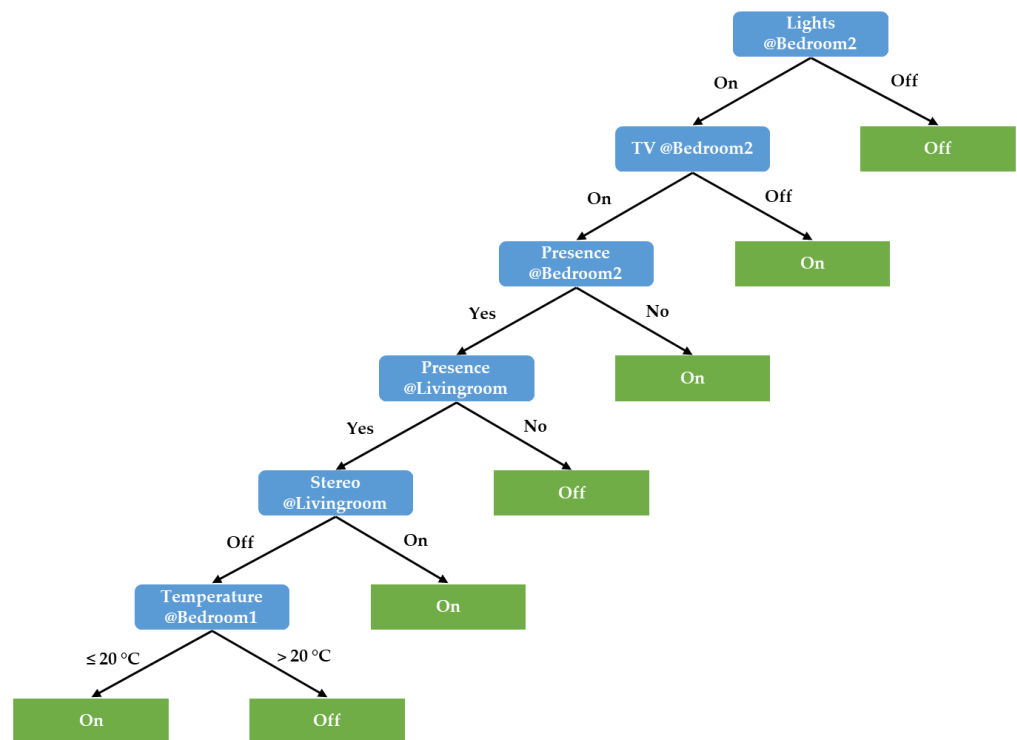


Figure 7. Resulting decision tree for living room lights.

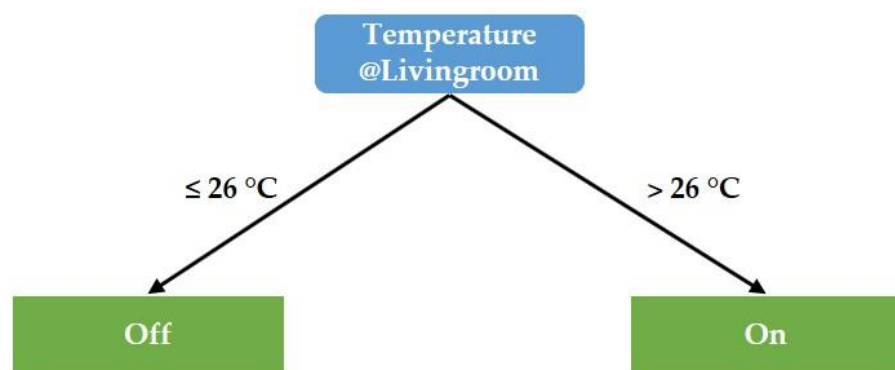


Figure 8. Resulting decision tree for living room air conditioner.

#### 4. Results and Discussion

As the case study has shown, through monitoring, collecting and analyzing the usage data coming from the devices connected to a home automation system, it was confirmed that the combination of machine learning and IoT allows identifying usage patterns. The usage patterns encoded as rules are of utmost importance as their application leads to improved comfort experience in smart homes and buildings. It should be noted that, although Smart Home Control has an intelligent control module with the ability to make autonomous decisions based on the usage history, the user retains the final authority over the control of home automation devices. By having the ability to record new changes in the state of home devices, the user will continue to record changes in the history of stored data, information that will serve to train Smart Home Control permanently and adaptively, allowing the decision trees to be readjusted as user behavior change.

In addition to the case study’s observation period described above, a second observation was carried out, for another ten months, from January to October 2020. The purpose of this second observation was to verify if the Automatic Control module of the Smart Home

Control system was capable of adjusting the automatic configuration rules for each device according to the changes in the new residents' usage patterns. It is important to mention that, although the Automatic Control was kept on during the second observation period, the ability of residents to control their devices manually was not limited. At the end of the second period, the following observations were established:

- The Automatic Control Module developed automatic configurations for each device, according to the data analysis carried out during the first observation period (March–December 2019).
- Residents could always cancel or update the orders established by the automatic control module to home automation devices.
- The devices' usage pattern of the residents is not necessarily static, as it also depends on external factors and unrelated to measurement devices, such as mood, health or other personal considerations.
- Significant social changes, such as the confinement measures established against the recent COVID-19 pandemic, drastically impacted residents' style and pace of life. This disruptive situation caused a radical change in the behavior patterns registered at home, which exacerbated the need for automatic configuration schemes.
- Some home automation devices may be less susceptible to subjective control factors by a resident. Consequently, these devices' control rules do not change—for example, the turning on or off of the odor extractor when a resident enters the bathroom.

Figures 9 and 10 show an example of the decision trees corresponding to the Bedroom 1 air conditioning device during Periods 1 and 2. A readjustment can be observed of the conditions considered for turning the device on and off. In turn, it led to a readjustment of the corresponding control rule due to changes in the behavior patterns of residents caused by the change of season of the year or confinement measures against the COVID-19 pandemic. One of the main differences between the two experiments is the considerable increase in house inhabitant presence registrations during the second period contemplated in 2020.

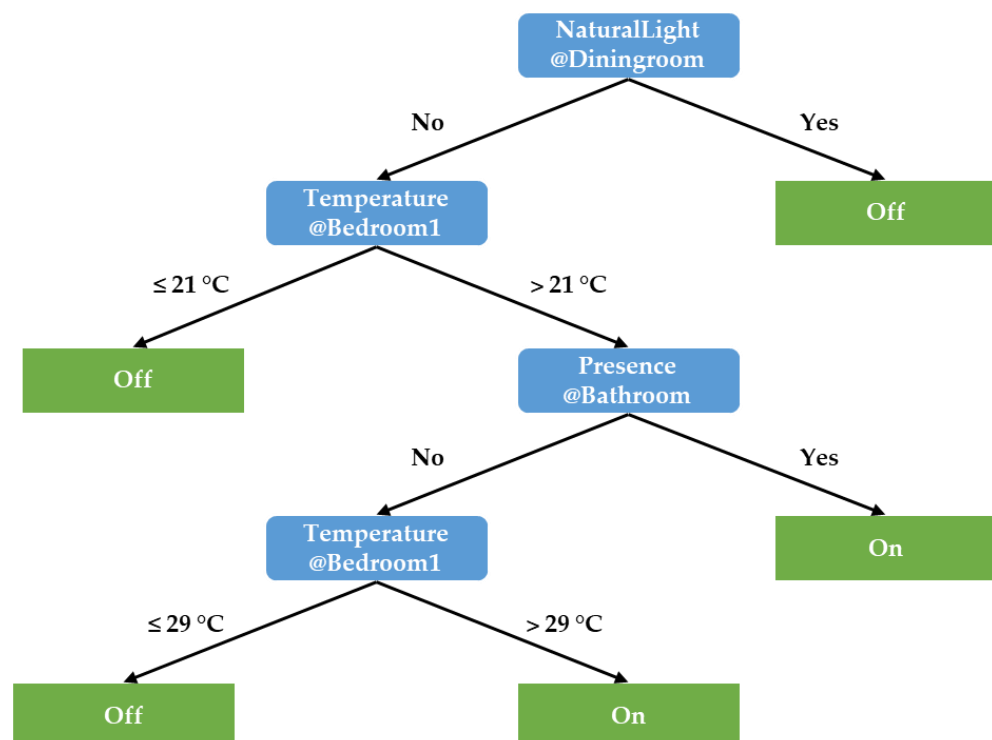


Figure 9. Decision tree for Bedroom 1 AC over Period 1.



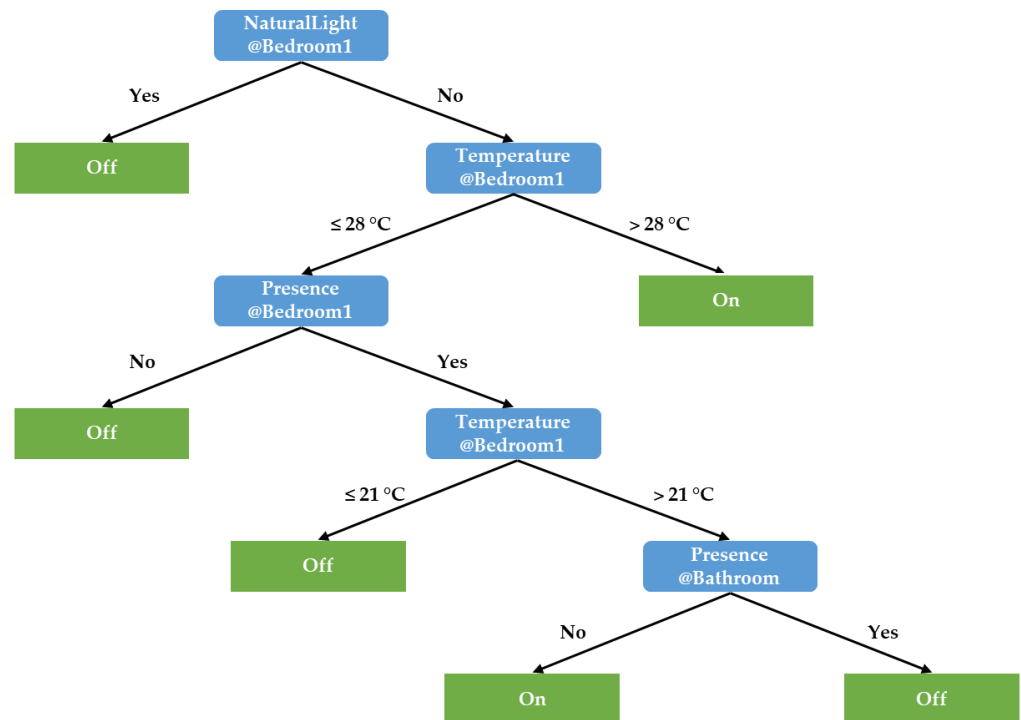


Figure 10. Decision tree for Bedroom 1 AC over Period 2.

Figures 11 and 12 show an example of the decision trees corresponding to the Bathroom odor extractor device during Periods 1 and 2, where no changes in the usage pattern were found.

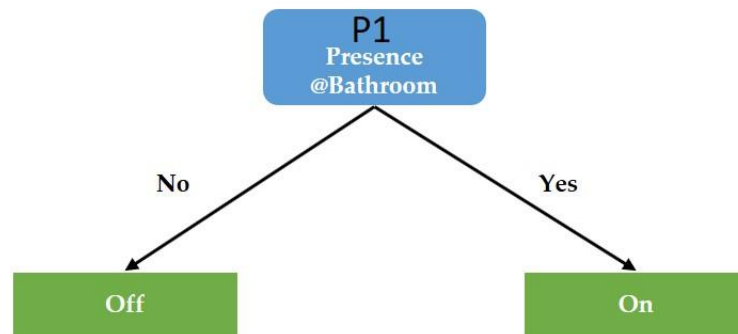


Figure 11. Decision tree for Bathroom Odor extractor over Period 1.

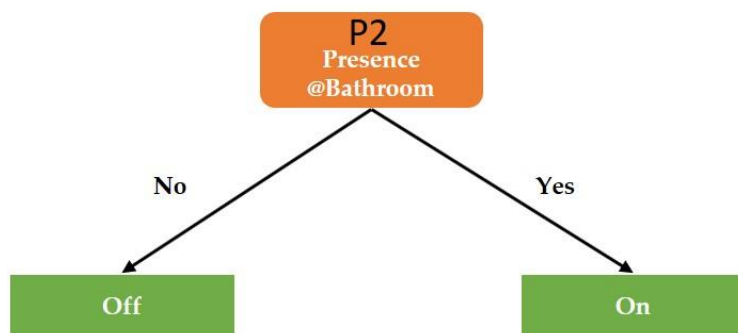


Figure 12. Decision tree for Bathroom Odor extractor over Period 2.

As verified, the automatic control module can readjust the home automation control rules according to the change in the residents' patterns of use. The readjust is possible thanks to the continuous recording and analysis of usage data from the devices connected to Smart Home Control, which allows the system to follow an "always on training" approach. This feature is of great importance if the automatic control module is required to make timely decisions considering its available information.

Throughout the development of the case study, it was possible to confirm that:

- The discovery of behavior patterns of the residents of a home is relevant to improving home automation conditions.
- The construction of automatic configuration rules must be dynamic and must evolve as residents' behavior patterns change.
- The analysis of the residents' behavior patterns accounts for continued training of the system. The training frees users from getting involved in making control decisions of their devices. However, no matter how close the system behaves to the residents, there will always be factors or conditions that cause changes in the home automation device usage.
- The automation of homes based on residents' behavior patterns brings benefits and improvements in home automation control schemes. Some of the schemes that can be improved under this approach are comfort, energy-saving, safety and healthcare schemes.

However, despite the benefits of developing automatic control of home automation devices based on the history of each house, there is a great challenge that must be taken into account:

- The human being is unpredictable, and behavior patterns can change drastically according to personal considerations, mood, mental health and medication.

Keeping in mind the volatility with which human beings change their behavior, it would be interesting to include technologies related to the reading and interpretation of body language to improve automatic configuration schemes.

Finally, it is worth mentioning that an extended analysis was performed despite obtaining satisfactory results using the C4.5 algorithm. The analysis aimed to determine, taking advantage of the information collected from the experimental use case, the feasibility of using other decision trees (e.g., CART) or ensemble algorithms (e.g., Adaboost and Random Forest). The analysis was conducted in the short term to improve attribute classification precision to construct automatic control rules. Table 6 shows the comparison of the precision metrics resulting from the analysis of attributes through the use of C4.5, CART, Adaboost and Random Forest.

**Table 6.** Algorithms performance metrics comparison.

Class Attribute	Algorithm	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
TV @Livingroom	C4.5	0.742	0.747	0.743	0.44	0.785	0.78
	CART	0.829	0.828	0.829	0.631	0.776	0.765
	Adaboost	<b>0.848</b>	<b>0.848</b>	<b>0.848</b>	<b>0.671</b>	0.883	0.854
	Random Forest	0.838	0.838	0.835	0.644	<b>0.902</b>	<b>0.899</b>
Stereo @Livingroom	C4.5	0.791	0.788	0.79	<b>0.396</b>	0.82	0.837
	CART	0.758	0.778	0.764	0.293	0.777	0.813
	Adaboost	0.71	0.747	0.722	0.152	0.811	0.853
	Random Forest	<b>0.802</b>	<b>0.818</b>	<b>0.791</b>	0.385	<b>0.859</b>	<b>0.874</b>

Table 6. Cont.

Class Attribute	Algorithm	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Lights @Livingroom	C4.5	0.833	0.848	0.838	0.405	0.604	0.782
	CART	0.722	0.798	0.75	0.017	0.486	0.705
	Adaboost	0.768	0.798	0.78	0.18	0.747	0.845
	Random Forest	<b>0.865</b>	<b>0.869</b>	<b>0.841</b>	<b>0.446</b>	<b>0.764</b>	<b>0.852</b>
AC @Livingroom	C4.5	<b>0.947</b>	<b>0.949</b>	<b>0.946</b>	<b>0.744</b>	0.722	0.883
	CART	0.912	0.919	0.912	0.571	0.733	0.874
	Adaboost	0.907	0.899	0.902	0.56	0.819	0.914
	Random Forest	0.854	0.879	0.861	0.295	<b>0.855</b>	<b>0.914</b>
Lights @Diningroom	C4.5	<b>0.876</b>	<b>0.879</b>	<b>0.877</b>	<b>0.639</b>	0.844	0.863
	CART	0.874	0.879	0.874	0.63	0.744	0.795
	Adaboost	0.862	0.869	0.862	0.594	0.894	0.919
	Random Forest	0.862	0.869	0.862	0.594	<b>0.94</b>	<b>0.944</b>
Lights @Kitchen	C4.5	0.819	0.808	0.813	0.456	0.787	0.826
	CART	<b>0.891</b>	<b>0.889</b>	<b>0.89</b>	<b>0.674</b>	0.744	0.796
	Adaboost	0.875	0.879	0.877	0.626	<b>0.921</b>	<b>0.926</b>
	Random Forest	0.793	0.808	0.798	0.376	0.897	0.891
Lights @Bathroom	C4.5	0.887	0.889	0.888	0.674	0.878	0.893
	CART	0.864	0.869	0.865	0.603	<b>0.907</b>	0.91
	Adaboost	0.821	0.828	0.824	0.48	0.876	0.902
	Random Forest	<b>0.927</b>	<b>0.919</b>	<b>0.912</b>	<b>0.759</b>	0.904	<b>0.938</b>
Extractor @Bathroom	C4.5	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.978</b>	0.986	0.977
	CART	0.96	0.96	0.959	0.909	0.981	0.969
	Adaboost	0.99	0.99	0.99	0.978	0.986	0.969
	Random Forest	0.98	0.98	0.98	0.955	<b>0.99</b>	<b>0.988</b>
TV @Bedroom1	C4.5	0.806	0.818	0.81	0.394	0.674	0.77
	CART	0.764	0.798	0.77	0.249	0.615	0.73
	Adaboost	0.786	0.778	0.782	0.336	0.778	0.836
	Random Forest	<b>0.852</b>	<b>0.859</b>	<b>0.839</b>	<b>0.499</b>	<b>0.857</b>	<b>0.904</b>
Lights @Bedroom1	C4.5	0.847	0.828	0.835	0.502	0.706	0.794
	CART	0.811	0.818	0.814	0.39	0.675	0.792
	Adaboost	<b>0.884</b>	<b>0.889</b>	<b>0.885</b>	<b>0.622</b>	0.875	0.886
	Random Forest	0.883	0.889	0.882	0.612	<b>0.913</b>	<b>0.936</b>
Lamps @Bedroom1	C4.5	0.818	0.818	0.818	0.329	0.616	0.767
	CART	0.793	0.808	0.8	0.236	0.703	0.8
	Adaboost	0.852	0.859	0.855	0.452	0.906	0.914
	Random Forest	<b>0.892</b>	<b>0.899</b>	<b>0.889</b>	<b>0.581</b>	<b>0.919</b>	<b>0.943</b>
AC @Bedroom1	C4.5	0.879	0.909	0.89	0.164	0.65	0.88
	CART	?	0.919	?	?	0.396	0.837
	Adaboost	0.879	0.909	0.89	0.164	0.674	0.894
	Random Forest	<b>0.934</b>	<b>0.929</b>	<b>0.903</b>	<b>0.341</b>	<b>0.821</b>	<b>0.937</b>
TV @Bedroom2	C4.5	0.871	0.869	0.859	0.652	0.791	0.833
	CART	<b>0.877</b>	<b>0.879</b>	<b>0.874</b>	<b>0.681</b>	0.755	0.78
	Adaboost	0.833	0.838	0.832	0.57	0.841	0.86
	Random Forest	0.833	0.838	0.634	0.576	<b>0.911</b>	<b>0.906</b>
Lights @Bedroom2	C4.5	<b>0.911</b>	<b>0.909</b>	<b>0.91</b>	<b>0.733</b>	0.933	0.921
	CART	0.854	0.859	0.856	0.563	0.793	0.846
	Adaboost	0.867	0.869	0.868	0.6	<b>0.955</b>	<b>0.956</b>
	Random Forest	0.906	0.909	0.906	0.716	0.931	0.94
Lamps @Bedroom2	C4.5	<b>0.835</b>	<b>0.838</b>	<b>0.837</b>	<b>0.594</b>	0.789	0.789
	CART	0.803	0.808	0.805	0.513	0.752	0.768
	Adaboost	0.818	0.818	0.818	0.552	0.875	0.898
	Random Forest	0.822	0.828	0.823	0.557	<b>0.908</b>	<b>0.924</b>

As can be seen, the use of Random Forest represents an improvement in the classification of attributes of the experimental case study. Table 7 shows the percentage of incidents in which each algorithm obtained a better score applicable to the attributes of the study case.

**Table 7.** Percentage of incidents in which each algorithm obtained a better score applicable to the attributes of the study case.

Algorithm	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
<b>C4.5</b>	33%	33%	27%	<b>40%</b>	0%	0%
<b>CART</b>	13%	13%	20%	13%	7%	0%
<b>Adaboost</b>	13%	13%	13%	13%	13%	13%
<b>Random Forest</b>	<b>40%</b>	<b>40%</b>	<b>40%</b>	33%	<b>80%</b>	<b>87%</b>

However, Random Forest's Smart Home Control implementation conveys challenging problems despite obtaining better scores in performance metrics. These problems are due to the difficulty interpreting the results obtained by the algorithm. These results contrast to those obtained by C4.5 whose iterative pruning process allows directly obtaining decision trees that are easily translated into control rules.

## 5. Conclusions

Humans have always been looking for more and better ways to facilitate daily life activities. It is evident that this desire also motivated the search to improve comfort conditions at home. Technologies such as machine learning and IoT make it possible to strengthen home automation systems with automatic control modules for daily use devices. It is important to mention that technological advances have allowed exponential growth in the number of IoT devices, which increases the ability to obtain information about the use that residents of a house give to their devices. Furthermore, it is possible to analyze the data coming from these IoT devices by using machine learning algorithms to identify residents' patterns of behavior to create automatic configuration schemes that adjust to the house residents' particular preferences.

In this paper, Smart Home Control is proposed, a platform that performs the analysis of historical records of the use of home automation devices to detect smart home residents' behavior patterns through IoT and machine learning, which improves the comfort schemes of domotic systems. Smart Home Control uses the C4.5 algorithm to classify the data from the sensors and IoT devices and thus build decision trees for each device's automatic configuration whose control adjusts to user's preferences. Additionally, an experimental case study was executed to validate the platform effectiveness. However, it should be noted that the results and effectiveness of Smart Home Control are affected by external factors such as social changes, natural disasters or even the mood of the inhabitants of the home. It is also important to emphasize that, after the implementation and results obtained using C4.5, it was possible to determine that the resulting models have a high level of interpretability. However, it was possible to identify a lack of explainability in the generated models as a limitation of this work. New evaluations are planned in the short term to establish strategies to deal with the black box problem generated by the low level of explainability.

Smart Home Control has high scalability potential. Therefore, future work is considered to incorporate more automatic configuration profiles related to security, energy-saving and accident detection, among others. With the increase in configuration profiles available to users, it will be possible to select the rules best suited for their needs, based on their current context. To validate the effectiveness of the new Smart Home Control profiles, it is planned to observe at least ten houses that make use of the profiles during monitoring periods of 8–10 months. The Smart Home Control mobile application was developed exclusively for Android, so it is desirable be available for other platforms. Regarding the

platform's effectiveness, it is important to record user satisfaction in terms of usability and improvement of their perception of comfort at home by applying user-centered evaluations based on the User-Centered Evaluation Framework for Computer Recommendation Systems. Finally, the implementation of more robust and complex classification algorithms, such as Random Forest, is considered for a possible improvement in the performance of the prediction process of the automatic control of home automation devices. In addition, according to the amount of data that are collected over time, the evolution of the data analysis paradigm of Smart Home Control to a Deep Learning and Big Data approach is feasible.

**Author Contributions:** Conceptualization, J.R.-C., G.A.-H. and I.M.-C.; Data curation, I.M.-C.; Formal analysis, L.R.-M. and J.O.O.-A.; Funding acquisition, G.A.-H., I.M.-C. and J.L.S.-C.; Investigation, J.L.S.-C.; Methodology, J.R.-C. and I.M.-C.; Project administration, G.A.-H.; Resources, G.A.-H.; Software, J.R.-C., I.M.-C. and L.R.-M.; Supervision, L.R.-M. and J.O.O.-A.; Validation, J.R.-C., G.A.-H. and L.R.-M.; Visualization, J.L.S.-C. and J.O.O.-A.; Writing—original draft, J.R.-C.; and Writing—review and editing, G.A.-H. and I.M.-C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Mexico's National Council of Science and Technology (CONACYT) through project 52–2016: "Application of Big Data and Semantic Web techniques to Develop Intelligent Systems", a postdoctoral grant, and a doctoral grant.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy concerns of the users involved in the study.

**Acknowledgments:** This work was supported by Mexico's National Technological Institute (TecNM) and sponsored by both Mexico's National Council of Science and Technology (CONACYT) and the Secretariat of Public Education (SEP) through the PRODEP project (Programa para el Desarrollo Profesional Docente).

**Conflicts of Interest:** The authors declare that there are no potential conflicts of interest with respect to the publication of this article.

## References

1. Reilly, E.D.; Ralston, A.; Hemmendinger, D. *Encyclopedia of Computer Science*; Nature Pub. Group: London, UK, 2000.
2. Thema, J.; Suerkemper, F.; Couder, J.; Mzavanadze, N.; Chatterjee, S.; Teubler, J.; Thomas, S.; Ürges-Vorsatz, D.; Hansen, M.B.; Bouzarovski, S.; et al. The Multiple Benefits of the 2030 EU Energy Efficiency Potential. *Energies* **2019**, *12*, 2798. [[CrossRef](#)]
3. Krishna, A.; Le Pallec, M.; Mateescu, R.; Noirie, L.; Salaun, G. IoT Composer: Composition and Deployment of IoT Applications. In Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), Montreal, QC, Canada, 25–31 May 2019; pp. 19–22.
4. Filho, G.P.R.; Villas, L.A.; Gonçalves, V.P.; Pessin, G.; Loureiro, A.A.; Ueyama, J. Energy-efficient smart home systems: Infrastructure and decision-making process. *Internet Things* **2019**, *5*, 153–167. [[CrossRef](#)]
5. Reyes-Campos, J.; Alor-Hernández, G.; Machorro-Cano, I.; Sánchez-Cervantes, J.L.; Muñoz-Contreras, H.; Olmedo-Aguirre, J.O. Energy Saving by Using Internet of Things Paradigm and Machine Learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media: Berlin/Heidelberg, Germany, 2020; pp. 447–458.
6. Paredes-Valverde, M.A.; Alor-Hernández, G.; García-Alcaráz, J.L.; Salas-Zárate, M.D.P.; Colombo-Mendoza, L.O.; Sánchez-Cervantes, J.L. IntelliHome: An internet of things-based system for electrical energy saving in smart home environment. *Comput. Intell.* **2020**, *36*, 203–224. [[CrossRef](#)]
7. Machorro-Cano, I.; Alor-Hernández, G.; Paredes-Valverde, M.A.; Rodríguez-Mazahua, L.; Sánchez-Cervantes, J.L.; Olmedo-Aguirre, J.O. HEMS-IoT: A Big Data and Machine Learning-Based Smart Home System for Energy Saving. *Energies* **2020**, *13*, 1097. [[CrossRef](#)]
8. Azar, J.; Makhoul, A.; Barhamgi, M.; Couturier, R. An energy efficient IoT data compression approach for edge machine learning. *Future Gener. Comput. Syst.* **2019**, *96*, 168–175. [[CrossRef](#)]



9. Valerio, L.; Conti, M.; Passarella, A. Energy efficient distributed analytics at the edge of the network for IoT environments. *Pervasive Mob. Comput.* **2018**, *51*, 27–42. [[CrossRef](#)]
10. Mahdavejad, M.S.; Rezvan, M.; Barekatin, M.; Adibi, P.; Barnaghi, P.; Sheth, A. Machine learning for internet of things data analysis: A survey. *Digit. Commun. Netw.* **2018**, *4*, 161–175. [[CrossRef](#)]
11. Filho, G.P.R.; Mano, L.Y.; Valejo, A.D.B.; Villas, L.A.; Ueyama, J. A Low-Cost Smart Home Automation to Enhance Decision-Making based on Fog Computing and Computational Intelligence. *IEEE Lat. Am. Trans.* **2018**, *16*, 186–191. [[CrossRef](#)]
12. Kasnesis, P.; Patrikakis, C.Z.; Venieris, I.S. Collective domotic intelligence through dynamic injection of semantic rules. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 592–597.
13. 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. [[CrossRef](#)]
14. Frontoni, E.; Liciotti, D.; Paolanti, M.; Pollini, R.; Zingaretti, P. Design of an interoperable framework with domotic sensors network integration. In Proceedings of the 2017 IEEE 7th International Conference on Consumer Electronics—Berlin (ICCE-Berlin), Berlin, Germany, 3–6 September 2017; pp. 49–50.
15. Silva, E.M.; Agostinho, C.; Jardim-Goncalves, R. A multi-criteria decision model for the selection of a more suitable Internet-of-Things device. In Proceedings of the 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC), Funchal, Portugal, 27–29 June 2017; pp. 1268–1276.
16. Wen, Y.; Wang, Z.; Zhang, Y.; Liu, J.; Cao, B.; Chen, J. Energy and cost aware scheduling with batch processing for instance-intensive IoT workflows in clouds. *Future Gener. Comput. Syst.* **2019**, *101*, 39–50. [[CrossRef](#)]
17. Terroso-Saenz, F.; González-Vidal, A.; Ramallo-González, A.P.; Skarmeta, A.F. An open IoT platform for the management and analysis of energy data. *Future Gener. Comput. Syst.* **2019**, *92*, 1066–1079. [[CrossRef](#)]
18. Pawar, P. Design and development of advanced smart energy management system integrated with IoT framework in smart grid environment. *J. Energy Storage* **2019**, *25*, 100846. [[CrossRef](#)]
19. Zekić-Sušac, M.; Mitrović, S.; Has, A. Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. *Int. J. Inf. Manag.* **2020**, 102074. [[CrossRef](#)]
20. Rafsanjani, H.N.; Ghahramani, A.; Nabizadeh, A.H. iSEA: IoT-based smartphone energy assistant for prompting energy-aware behaviors in commercial buildings. *Appl. Energy* **2020**, *266*, 114892. [[CrossRef](#)]
21. Martínez, E.G.-; Rodrigues, C.F.; Rileyb, G.; Grahn, H. Estimation of energy consumption in machine learning. *J. Parallel Distrib. Comput.* **2019**, *134*, 75–88. [[CrossRef](#)]
22. Li, X.; Chen, S.-Q.; Li, H.; Lou, Y.; Li, J. Multi-dimensional analysis of air-conditioning energy use for energy-saving management in university teaching buildings. *Build. Environ.* **2020**, *185*, 107246. [[CrossRef](#)]
23. Raza, R.; Hassan, N.U.; Morstyn, T. Determination of consumer behavior based energy wastage using IoT and machine learning. *Energy Build.* **2020**, *220*, 110060. [[CrossRef](#)]
24. Chacon-Troya, D.P.; Gonzalez, O.O.; Campoverde, P.C. Domotic application for the monitoring and control of residential electrical loads. In Proceedings of the 2017 IEEE 37th Central America and Panama Convention (CONCAPAN XXXVII), Managua, Nicaragua, 15–17 November 2017; pp. 1–6.
25. Saba, D.; Degha, H.E.; Berbaoui, B.; Laallam, F.Z.; Maouedj, R. Contribution to the modeling and simulation of multiagent systems for energy saving in the habitat. In Proceedings of the 2017 International Conference on Mathematics and Information Technology (ICMIT), Adrar, Algeria, 4–5 December 2017; pp. 204–208. [[CrossRef](#)]
26. Buono, P.; Balducci, F.; Cassano, F.; Piccinno, A. EnergyAware: A non-intrusive load monitoring system to improve the domestic energy consumption awareness. In *EnSEmble 2019: Proceedings of the 2nd ACM SIGSOFT International Workshop on Ensemble-Based Software Engineering for Modern Computing Platforms August 2019*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–8. [[CrossRef](#)]
27. Li, W.; Logenthiran, T.; Phan, V.-T.; Woo, W.L. Implemented IoT-Based Self-Learning Home Management System (SHMS) for Singapore. *IEEE Internet Things J.* **2018**, *5*, 2212–2219. [[CrossRef](#)]
28. Elkhorchani, H.; Grayaa, K. Novel home energy management system using wireless communication technologies for carbon emission reduction within a smart grid. *J. Clean. Prod.* **2016**, *135*, 950–962. [[CrossRef](#)]
29. Matsui, K. An Information Provision System as a Function of HEMS to Promote Energy Conservation and Maintain Indoor Comfort. *Energy Procedia* **2017**, *105*, 3213–3218. [[CrossRef](#)]
30. Al-Turjman, F.; Altrjman, C.; Din, S.; Paul, A. Energy monitoring in IoT-based ad hoc networks: An overview. *Comput. Electr. Eng.* **2019**, *76*, 133–142. [[CrossRef](#)]
31. Baker, T.; Asim, M.; Tawfik, H.; Aldawsari, B.; Buyya, R. An energy-aware service composition algorithm for multiple cloud-based IoT applications. *J. Netw. Comput. Appl.* **2017**, *89*, 96–108. [[CrossRef](#)]
32. Ganesan, T.; Vasant, P.; Elamvazuthi, I. *Advances in Metaheuristics: Applications in Engineering Systems*; CRC Press: Boca Raton, FL, USA, 2016.
33. Vasant, P.; Kose, U.; Watada, J. Metaheuristic Techniques in Enhancing the Efficiency and Performance of Thermo-Electric Cooling Devices. *Energies* **2017**, *10*, 1703. [[CrossRef](#)]
34. Zelinka, I.; Tomaszek, L.; Vasant, P.; Dao, T.T.; Hoang, D.V. A novel approach on evolutionary dynamics analysis—A progress report. *J. Comput. Sci.* **2018**, *25*, 437–445. [[CrossRef](#)]

35. Vasant, P.; Marmolejo, J.A.; Litvinchev, I.; Aguilar, R.R. Nature-inspired meta-heuristics approaches for charging plug-in hybrid electric vehicle. *Wirel. Netw.* **2019**, *26*, 4753–4766. [[CrossRef](#)]
36. Li, D.; Deng, L.; Lee, M.; Wang, H. IoT data feature extraction and intrusion detection system for smart cities based on deep migration learning. *Int. J. Inf. Manag.* **2019**, *49*, 533–545. [[CrossRef](#)]
37. Alli, A.A.; Alam, M.M. SecOFF-FCIoT: Machine learning based secure offloading in Fog-Cloud of things for smart city applications. *Internet Things* **2019**, *7*, 100070. [[CrossRef](#)]
38. Lanfor, O.G.F.; Perez, J.F.P. Implementación de un sistema de seguridad independiente y automatización de una residencia por medio del internet de las cosas. In Proceedings of the 2017 IEEE Central America and Panama Student Conference (CONESCAPAN), Panama City, Panama, 20–22 September 2017; pp. 1–5.
39. Malina, L.; Srivastava, G.; Dzurenda, P.; Hajny, J.; Fajdiak, R. A Secure Publish/Subscribe Protocol for Internet of Things. In *ARES '19, Proceedings of the 14th International Conference on Availability, Reliability and Security, Canterbury, UK, 26–29 August 2019*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–10.
40. Mozaffari, N.; Rezazadeh, J.; Farahbakhsh, R.; Yazdani, S.; Sandrasegaran, K. Practical fall detection based on IoT technologies: A survey. *Internet Things* **2019**, *8*, 100124. [[CrossRef](#)]
41. Din, I.U.; Guizani, M.; Rodrigues, J.J.; Hassan, S.; Korotaev, V. Machine learning in the Internet of Things: Designed techniques for smart cities. *Future Gener. Comput. Syst.* **2019**, *100*, 826–843. [[CrossRef](#)]
42. Nizetic, S.; Šolić, P.; González-De-Artaza, D.L.-D.-I.; Patrono, L. Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. *J. Clean. Prod.* **2020**, *274*, 122877. [[CrossRef](#)]
43. Singh, S.; Gupta, P. Comparative study id3, cart and c4.5 decision tree algorithm: A survey. *Int. J. Adv. Inf. Sci. Technol.* **2014**, *27*, 97–103.
44. Sood, S.K.; Mahajan, I. A Fog-Based Healthcare Framework for Chikungunya. *IEEE Internet Things J.* **2018**, *5*, 794–801. [[CrossRef](#)]
45. Kureshi, N.; Abidi, S.S.R.; Blouin, C. A Predictive Model for Personalized Therapeutic Interventions in Non-small Cell Lung Cancer. *IEEE J. Biomed. Health Inform.* **2016**, *20*, 424–431. [[CrossRef](#)] [[PubMed](#)]
46. Pintelas, E.; Livieris, I.E.; Pintelas, P. A Grey-Box Ensemble Model Exploiting Black-Box Accuracy and White-Box Intrinsic Interpretability. *Algorithms* **2020**, *13*, 17. [[CrossRef](#)]