



# Article Decision Support System to Classify and Optimize the Energy Efficiency in Smart Buildings: A Data Analytics Approach

Manuel Peña \*, Félix Biscarri, Enrique Personal 🗅 and Carlos León 🕒

Electronic Technology Department, School of Computer Science and Engineering, University of Seville, Av. Reina Mercedes S/N, 41012 Seville, Spain; fbiscarri@us.es (F.B.); epersonal@us.es (E.P.); cleon@us.es (C.L.) \* Correspondence: manuelpm@dte.us.es

Abstract: In this paper, an intelligent data analysis method for modeling and optimizing energy efficiency in smart buildings through Data Analytics (DA) is proposed. The objective of this proposal is to provide a Decision Support System (DSS) able to support experts in quantifying and optimizing energy efficiency in smart buildings, as well as reveal insights that support the detection of anomalous behaviors in early stages. Firstly, historical data and Energy Efficiency Indicators (EEIs) of the building are analyzed to extract the knowledge from behavioral patterns of historical data of the building. Then, using this knowledge, a classification method to compare days with different features, seasons and other characteristics is proposed. The resulting clusters are further analyzed, inferring key features to predict and quantify energy efficiency on days with similar features but with potentially different behaviors. Finally, the results reveal some insights able to highlight inefficiencies and correlate anomalous behaviors with EE in the smart building. The approach proposed in this work was tested on the BlueNet building and also integrated with Eugene, a commercial EE tool for optimizing energy consumption in smart buildings.

**Keywords:** smart building; energy efficiency; data analytics; energy optimization; decision support system

# 1. Introduction

The growth of energy consumption, energy resource exhaustion and significant environmental impacts [1,2] have raised concerns in most countries, which have entered international agreements for the benefit of society, such as the Paris Agreement in 2015 [3]. With a total energy consumption of 41% [4,5] in residential, public service and commercial sectors, which represents 24% of the world's  $CO_2$  emissions [6], these sectors constitute one of the key areas of interest to address, where most action plans are focused on improving Energy Efficiency (EE) through the promotion of renewable energies and also evolving systems to minimize energy consumption. Given this scenario, this work tackles the EE challenge by proposing a new approach to support these initiatives through the optimization of EE in smart buildings using data analytics techniques with the aim of reducing energy consumption and, therefore, reducing  $CO_2$  emissions.

Because of the relevance of these initiatives for the future of society, many international organizations are sponsoring these initiatives [7,8], such as the International Energy Agency (IEA), the U.S. Department of Energy—Energy Information Administration (DOE/EIA), the Energy European Commission (EEC) and the Organisation for Economic Co-operation and Development (OECD), with directives such as Energy Efficiency Directive 2012/27/EU (EU) or Executive Order (EO) 13514, Federal Leadership in Environmental, Energy, and Economic Performance, October 2009 (DOE/EIA). Additionally, substantial opportunities to improve energy efficiency have been expressed by the International Energy Agency (IEA), as described in [9]. These international initiatives, along with new technologies and their related research proposals, are actively contributing to tackling the above-mentioned



Citation: Peña, M.; Biscarri, F.; Personal, E.; León, C. Decision Support System to Classify and Optimize the Energy Efficiency in Smart Buildings: A Data Analytics Approach. *Sensors* 2022, 22, 1380. https://doi.org/10.3390/s22041380

Academic Editor: Hossam A. Gabbar

Received: 10 January 2022 Accepted: 27 January 2022 Published: 11 February 2022

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



**Copyright:** © 2022 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/). challenges. Accordingly, with rising concerns about the exhaustion of energy resources and significant environmental impacts around the world, EE consumption is a rapidly increasing field.

From this perspective, in the last decade, huge efforts have been dedicated to the development of smart cities [10–12] in an attempt to design more sustainable cities by applying green energies and a high level of smart services [13–21] with the support of new technologies such as IoT [22–26] and 5G [27,28] and the progress made in AI [29–37], transport [38], etc. Within this context, efforts have been dedicated to increasing EE in smart buildings with more intelligent behavior, a high level of comfort and environmentally friendly operation [34,37,39–44].

There is considerable research focus on energy optimization in buildings. The increasingly sophisticated Building Automation System (BAS) [45] has become the cornerstone of intelligent modern buildings, integrating energy supply and demand factors, often known as Demand-Side Management (DSM), as energy efficiency policies that predict demand, as presented in [46]. Accordingly, most of the progress has been made in Energy Management and Control Systems (EMCSs) [47], which can manage energy policies in real time with the aim of maintaining a high level of comfort with minimum power consumption in different operating conditions [18,48,49]. In EMCSs, one of the most complex problems is optimization according to real-time environmental variables of the building. Current research approaches try to solve this problem from different points of view and with diverse techniques [50] and in different potential scenarios [51-53]. These approaches started with basic actions to improve EE [51] until more sophisticated energy management models were introduced, such as those based on a predictive controller in a Supervisory Control and Data Acquisition (SCADA) system [54], multi-objective models through a multi-criteria decision analysis [55–57], a more complex approach with Multi-Agent Systems (MAS) [58] or MAS based on occupant behaviors [59], a model based on a Markov Decision Model (MDM) [60] or a predictive control model combined with weather forecasting [61].

Within the above context, the current work is part of the KnoHolEM project, associated with FP7 (FP7-285229 KnoHolEM). The main purposes of this project are mainly focused on achieving the following seven objectives: (1) a functional energy-oriented building model complemented by a corresponding generic building ontology, (2) a specific building behavior model completed by a building-specific ontology, (3) data-mining procedures for detailed real-time energy consumption analysis, (4) algorithms derived from the building-specific ontology running in real time to acquire energy efficiency measures, (5) software for the synthesis and validation of real-time control algorithms, (6) the definition and engineering of hardware and firmware for real-time communication and optimization of energy in buildings and (7) and an interactive virtual reality smart building simulator.

In this sense and following the third and fourth objectives of the KnoHolEM project, the aim of the present paper is to optimize EE in smart buildings through a data analytics approach based on applying Data-Mining (DM) techniques, reducing energy consumption, maintaining a high degree of comfort and being environmentally friendly. Several research efforts have been dedicated to studying energy optimization in buildings with different techniques [62]. The efficiency of a building depends heavily on the way that it is used and how it is managed. In [63], the authors provided an interesting analysis of energy-efficient building design through DM techniques. In [64], the semantic modeling of building systems to support advanced data analytics for EE improvements was described. The authors of [65] presented an advanced DA framework for EE in buildings. However, the analysis of energy consumption at the use stage is essential, and it is clear that the construction characteristics of buildings strongly affect their energy consumption during their life cycle. In [66], the authors presented a review of unsupervised analytics techniques applied to EE enhancement. Finally, [67] proposed a classification approach of energy consumption in buildings.

In this work, a system to classify and measure EE in a smart building is presented. This system is implemented after an analytical process is performed to extract the knowledge

hidden in historical building data. In addition, the systems are based on a set of models and algorithms corroborated by EE experts and historical data. This knowledge is used to classify each day analyzed based on its features to then predict the EE based on the insights extracted from historical data and experts. The purpose of this study is the optimization of EE based on measuring and predicting EE for each day from historical and real-time data of the BlueNet building. Subsequently, from the results obtained, it is easy to detect patterns that correlate days with poor EE with a high probability of anomalies. Finally, thanks to the progress made in this field and the results obtained, the system is in the test phase. A future goal is to integrate this framework with commercial EE software called Eugene, owned by the company Isotrol (Seville, Spain).

The paper is organized as follows: In Section 2, the smart building description and data sources of this work are presented. The description of data sources is divided into four subsections: indoor sensors, outdoor sensors, energy analyzers and, finally, EEIs. In Section 3, the proposed method, a data-mining-based DSS to measure and increase energy efficiency in the building, is detailed. It is divided into three subsections: data preprocessing (cleaning, integration, reduction and selection, and transformation), description of the classification module and, finally, the energy efficiency prediction module. In Section 4, the results of the work are presented.

#### 2. Case Study: BlueNet Smart Building

Usually, smart buildings are made up of several data-metering units that are managed by an EMCS. This EMCS is able to act as a powerful tool for increasing EE in the building. For EE optimization, it is necessary to know how patterns of energy consumption in the building have been behaving. This knowledge is essential for understanding the consumption patterns of the BlueNet building and, therefore, for optimizing the energy efficiency of the consumption process. Thus, in this section, the BlueNet building and the technology employed in the smart building are described in detail. In addition, a description of data sources used for the classification and prediction of EE in the building is presented.

The BlueNet building is situated in the south of Spain (Seville, 37°24'29.97" N, 6°00'18.63" W). This building is used as an office building, and its main activity occurs from 8:00 to 17:30, although there is some variability in the schedule. The data obtained from the BlueNet building are from several data sources, such as indoor sensors, outdoor sensors, energy analyzers and EEIs, covering the main four areas of the building. The indoor sensors are made up of 16 ZigBee sensors per floor that collect indoor variables; ZigBee motes are centralized by a master ZigBee mote used as a gateway that gathers all measurements and sends them to a centralized server. Occupancy data can be obtained by many approaches, such as Passive Infrared (PIR) sensors, ultrasonic systems, camera-based systems, radar systems,  $CO_2$  sensing, Electromagnetic (EM) signal detection systems, energy measurement devices, computer activity or sensor fusion, chair sensors or the use of multiple technologies to enhance the results [68,69]. In our case, the measurements were obtained through a combination between Radio Frequency Identity (RFID) technology, which is used by proximity cards at the entrance and exit of each room for exhaustive control for staff identity purposes, and Passive Infrared (PIR) sensors incorporated in Zigbee motes positioned strategically throughout the building space. Outdoor sensors comprise a set of sensors placed on the BlueNet building to extract environmental variables and send all of their measurements to a centralized server. Finally, the energy analyzers measure all power consumption variables of the building (HVAC, air mixers, splits, lighting, power plugs and other consumptions). These analyzers are centralized through a master Modbus device. The master Modbus device collects and transmits all consumption measurements to a centralized server that collects all of the BlueNet building information. Besides the information on these areas, a set of useful EEIs and information concerning the schedule of working and non-working days are shared, as this is quite valuable information to understand building consumption patterns and apply the knowledge extracted in this work.

#### 2.1. Indoor Sensors

These sets of ZigBee sensors are placed in strategic locations in the BlueNet building and coordinated by a master sensor. The master sensor collects the measurements of the other sensors and sends the data to a master device that acts as a gateway. The purpose of this gateway is to collect sensor information and send it to the BlueNet centralized server. Indoor sensors are programmed to send meter readings every 30 s. These sensors collect the following measurements:

- Temperature: ZigBee sensors read the temperature in every room of the building. Generally, at least 2 or 3 sensors are used in every room to apply rules that ensure that the correct temperature is obtained.
- Humidity: ZigBee sensors read the relative humidity in every room of the BlueNet building. It is important to calculate the real feeling of the temperature in each zone.
- Lux: ZigBee sensors check the real effect of the lighting system by measuring the lux values in every room of the building.
- Presence: These sensors obtain data that indicate the presence of all occupants in the BlueNet building by identifying every person with a unique id and obtaining information about the time that every person is in BlueNet building rooms.

## 2.2. Outdoor Sensors

In addition, there is a set of sensors that collect metering data of the BlueNet environment. These sensors are responsible for collecting environmental data. Outdoor sensors are programmed to measure metering data every 10 min. The main measurements are:

- Temperature: These sensors take the environmental temperature every 10 min and are also able to provide the maximum, minimum and mean temperature of each day.
- Humidity: The outdoor sensors measure the environmental humidity and the amount of rain fallen to calculate the feeling of environmental temperature.
- Sunshine: These sensors obtain the amount of sunshine that irradiates onto the building every day.
- Wind: The outdoor sensors also obtain the amount and direction of the wind in the building environment.

## 2.3. Energy Analyzers

Energy analyzers are placed in an energy distribution panel inside of the BlueNet building. They are connected to a Programmable Logic Controller (PLC), which acts as a master Modbus device of the energy analyzers. These analyzers meter the four typologies of energy consumption in the building (HVAC, lights, power plugs and other consumptions). These analyzers are programmed to send measurements every 5 min to the master Modbus device. The principal features of every typology of consumption are described below.

## 2.3.1. HVAC

In the BlueNet building, HVAC is based on a VRV system. This system is made up of a set of indoor units (Daikin FXSQ-M7V1B, Daikin AC Spain S.A, Madrid, Spain) and a set of outdoor units (Daikin RXYQ-MY1B VRV II inverter—Daikin AC Spain S.A, Madrid, Spain—with heat pump). All units are connected by a DIII-Net. These connections are centralized in a DMS504B51 Daikin Lonworks Interface, which is in turn connected to a centralized server through a Lonworks/Modbus gateway (IntesisBox—HMS Industrial Networks AB S.L.U., Barcelona, Spain) using the communication protocol I3E.

HVAC was the primary area of consumption in this study due to its strong influence on EE [50–52]. Specifically, HVAC systems are appliances with the largest consumption in the building and are also the most controllable. HVAC management has the largest margin for EE improvement, consuming 143,876.7 MWh in 509 days, which represents 40.11% of the total energy consumption in the building.

## 2.3.2. Lights

The BlueNet building is made up of a luminary system, Philips Light Master (Luminary with ballast HF-R TD 318), connected through a DALI bus. This bus is centralized and managed by an LRC5141 controller, and it is connected to a centralized server through a Lonworks/Modbus gateway (IntesisBox—HMS Industrial Networks AB S.L.U., Barcelona, Spain). This server is responsible for sharing the data and applying certain energy management policies.

The lighting only accounted for 19.03% of total energy consumption in the BlueNet building, consuming 68,281.7 MWh throughout the 509 studied days. Thus, it is difficult to reduce this consumption (lighting is a necessary building function, and it is associated with occupancy in rooms). Nonetheless, the study of this field provides interesting information on EE, such as occupation patterns, anomalous consumptions or group behaviors.

#### 2.3.3. Power

The consumption at power points in the BlueNet building is mainly the result of ICT equipment (PCs, servers, media, etc.). The majority of power consumption usually occurs during working hours, with the exception of some services that provide support 24 h a day.

This area is the second most relevant for this study, accounting for 120,368.7 MWh of the total consumption in the study period. Although it constitutes the second largest consumption in the building, accounting for 33.56% of the total, this energy consumption is hard to improve from the EE point of view because computers perform scheduled tasks outside of working hours, making it difficult to reduce this type of energy consumption.

# 2.3.4. Others

This area accounts for the minority of energy consumption, with only 26,170 MWh of total consumption. In addition, it has a strongly fixed consumption that is difficult to manage. Thus, with only 7.3% of total energy consumption, it is the least relevant field in EE.

## 2.4. Energy Efficiency Indicators (EEIs)

A set of EEIs of the BlueNet building were analyzed. These EEIs are based on experts' knowledge and historical data behavior of each relevant area in the BlueNet building, as described in [21]. The EE behavior in the BlueNet building is evaluated on the basis of these EEIs. In addition, each EEI provides the knowledge required to detect EE behaviors and anomalies. The EEIs are described further below.

# 2.4.1. Operational Changes in HVAC Compressor (OCC) Indicator

This EEI counts the number of daily on–off operations in the compressor. A large number of daily on–off operations are considered anomalous or inefficient, and it can cause one of the largest energy leaks and high inefficiency, greatly increasing energy consumption. A high on–off operation variance could indicate a possible anomaly in HVAC management (the HVAC is poorly dimensioned for this room, the HVAC is too powerful for this room, or there is a possible malfunction in the temperature sensor). Moreover, a compressor with a high OCC is prone to break down and have a shorter lifetime.

#### 2.4.2. Number of Operational Regime Changes in the HVAC Compressor (ORCC) Indicator

This EEI counts the number of daily ORCC periods and the number of minutes in which the daily ORCC periods occurred. An ORCC is defined as a change in the compressor power consumption greater than 0.5 kW with respect to the previous measurement (10 min). These parameters were specified based on the results of DM techniques under the consensus of Isotrol HVAC experts. Thus, a large number of operational regime changes in the HVAC compressor (ORCC) is considered abnormal or inefficient.

This EEI can indicate that the HVAC system is not properly calibrated (HVAC is too powerful for this requirement) or that the temperature in the room is forcing the HVAC compressor to constantly change its operating mode. Furthermore, this EEI can denote a possible improvement in HVAC management, softening the HVAC consumption curve.

2.4.3. Switch on HVAC Compressor and Abnormal Changes in Indoor Temperature (SONCCIT) Indicator

This EEI counts the number of total minutes per day with a SONCCIT anomaly. In addition, this EEI averages the daily active power in the HVAC compressor during the compressor anomaly and the number of periods per day in which a switched-on compressor anomaly is observed (an anomalous period is specified as the aggregation of consecutive anomalous data points). A SONCCIT anomaly is defined when the HVAC system is turned on (the HVAC compressor consumption is higher than 1.7 kWh) and produces a change in indoor temperature greater than or equal to 1 °C between samples (10 min).

It is considered abnormal or inefficient when the room temperature decreases sharply (winter) or increases sharply (summer) while the compressor is running. This could be due to a sudden leakage of heat (winter) or a sudden influx of heat (summer) in the room, which counteracts the effect of the HVAC system (i.e., opened windows or doors).

2.4.4. Switch off HVAC Compressor and Abnormal Changes in Indoor Temperature (SOFFCCIT) Indicator

This EEI counts the total minutes per day with a SOFFCCIT anomaly. In addition, this EEI averages the daily active power in the HVAC compressor and the number of periods per day in which a switched-off compressor anomaly is observed (as in SONCCIT, an anomalous period is defined as the aggregation of consecutive anomalous data points).

It is considered abnormal or inefficient when the room temperature rises sharply (winter) while the compressor is not running. This could be due to a heat source (electric heater) replacing the HVAC system and can indicate the inefficient use of power energy. From historical data, a 1 °C increase between samples (10 min) during the winter period or a 1 °C decrease during the summer months is considered anomalous.

#### 2.4.5. No Persons in BlueNet Building and Switch on HVAC Compressors (NPSONC) Indicator

This EEI counts the total daily minutes in which an NPSONC anomaly is detected, the number of periods per day with an NPSONC anomaly (an anomalous period is defined as the aggregation of consecutive anomalous data points) and the average active power consumption per day by the HVAC compressor during the anomaly. An anomalous function of the compressor is identified when there is an absence of occupants or the lights are switched off and the compressor is switched on (NPSONC). This could indicate that the air conditioner is switched on accidentally, considering that there are no building occupants if the lights are not switched on.

In this study, five data sources were analyzed: indoor sensors, outdoor sensors, schedule of working days, energy analyzers and EEIs. These data sources were employed during this work, in which every area was exhaustively and carefully examined to detect any possible improvement in energy management.

Some important information for this first analysis of historical data behaviors yielded the following results: 40.11% of building consumption was attributable to HVAC, 33.56% of energy consumption was due to power, 19.03% was spent on lighting, and another 7.3% was due to other activities, with an accumulated energy consumption of 358,697.1 MWh in the BlueNet building during the analyzed period. The most relevant consumption of the BlueNet building was due to HVAC operation with 143,876.7 MWh, and specifically, the major HVAC consumption was attributable to the compressor motor engine consumption, accounting for 46.38% of the total HVAC consumption.

Thus, a system to measure and optimize EE in the BlueNet building through DA techniques was developed. The objectives of this system were to: extract the knowledge hidden in BlueNet building data, develop a classification module and build an EE prediction module that helps to predict EE for each day. This EE classification is able to compare EE on days with similar characteristics, regardless of the season and other factors, which

would be difficult to compare without this information, and also allows the presence of anomalous energy consumptions and other possible problems to be identified. The EE prediction module is able to quantify the energy efficiency every day, comparing days with similar energy efficiency conditions based on clusters.

The architecture of the EE optimization system is depicted in Figure 1.



Figure 1. Prototype architecture.

## 3. The Data-Mining-Based Decision Support System to Optimize EE in the Smart Building

This section describes the approach selected to optimize the EE in the BlueNet building. This approach is based on a hybrid decision support system that combines a search based on the historical data to classify the energy consumption on days with similar features and a prediction module for additional EE interpolations. This approach takes advantage of the knowledge extracted from the historical data of the building, making it possible to reveal information about behavioral patterns in energy management, as well as knowledge provided by HVAC experts to optimize EE. Therefore, this approach comprises a module for the classification of energy consumption, which is based on CR&T decision tree [70] and cluster classification, and a module for EE prediction, which is based on metrics. With both modules, the days are analyzed with the EEI results to discover patterns and detect anomalies or other possible problems.

Data can provide some insights hidden in the behavior of historical data. Normally, the expert's knowledge is hidden in the collected dataset. Knowledge Discovery in Databases (KDD) refers to the overall data-mining process of discovering useful knowledge from large amounts of data. The DM process consisted of 6 essential phases: understanding the business, understanding the data, data preprocessing, modeling, evaluation and deployment. Once the phases of understanding the business and data were carried out, the next phase was the preprocessing of data. In this phase, the data were cleaned, and the different data sources were integrated, reduced and selected, and finally, transformed [71,72] (described in Section 3.1). After these phases, data were prepared for the modeling phase, which is explained in Sections 3.2 and 3.3 and evaluated in Section 4.

On the one hand, the first study aimed to extract the knowledge of energy consumption behaviors of the BlueNet building and determine how the energy has been consumed (e.g., regime changes in the compressor, operational changes in the compressor, lighting patterns and others). Secondly, the influence of each BlueNet building variable in every area (indoor sensors, outdoor sensors, HVAC, lighting, power and other consumption measurements, and EEIs) was quantified.

On the other hand, after extracting the knowledge hidden in BlueNet building data, an EE classification of the historical data was carried out through a hybrid system of classification. This EE classification allows the establishment of a relationship between days with similar features and also enables experts to compare the behavior on these days in EE terms. The classification module is based on a decision tree, the clustering of historical building data and the use of a set of energy efficiency indicators (Section 3.2). As a result, this classification identifies the presence of anomalous energy consumptions and other possible problems. In addition, a prediction system to quantify EE every day was developed (Section 3.3). This prediction is based on the features of each cluster classification, with the aim of assisting in the quantification of energy efficiency every day.

In order to carry out this work, a powerful and extended tool in the analytics area was used, namely, the SPSS Modeler tool (originally Statistical Package for Social Sciences Inc., an IBM company). This tool was used to perform the preprocessing and modeling tasks, as well as to evaluate the models. In addition, SPSS Modeler includes IA libraries used by the classification and prediction systems through DM techniques.

Currently, the decision support system to optimize EE in the BlueNet building is in the testing phase, and it will be connected to Eugene, an EE tool owned by Isotrol. These modules provided all of the knowledge extracted from the large amount of data provided by the BlueNet building with the aim of improving the results of the EE tool.

#### 3.1. Data Preprocessing

The DM process requires an initial phase of data preprocessing, in which the data are analyzed, filtered and formatted [71]. BlueNet data comprise different data sources: energy consumption, environmental sensors inside of the building, external climate sensors, EEIs and other sources of data, all of which have their own temporal frequency. Thus, to manage different timestamps among the recorded data, the frequency was synchronized with a period of 10 min. The time interval for these data sources is between January 2011 and March 2013 (509 days).

The type of data strongly depends on the source of the data. The types of data and their time bases are as follows:

- Indoor sensors (30 s basis): mote\_id, timestamp (YYYY/MM/DD hh:dd:ss), temperature (Celsius degrees), percentage\_humidity, CO<sub>2</sub> and lux (lumens).
- Indoor sensors (30 s basis): mote\_id, timestamp (YYYY/MM/DD hh:dd:ss) and employee\_id (presence).
- Outdoor sensors (10 min basis): sensor\_id, latitude, longitude, timestamp (yyyy/mm/dd hh:dd:ss), wind\_direction (degrees), max\_wind\_speed (m/s), min\_wind\_speed (m/s), ave\_wind\_speed (m/s), UV\_index, max\_humidity, min\_humidity, ave\_humidity, precipitation (l/m<sup>2</sup>) and sunshine\_radiation (W/m<sup>2</sup>).
- Energy Analyzers—HVAC (5 min basis): timestamp (YYYY/MM/DD hh:mm:ss), AP\_CLI\_FASE1 (kW), AP\_CLI\_FASE2 (kW) and AP\_CLI\_FASE3 (kW).
- Energy Analyzers—Lights (5 min basis): timestamp (YYYY/MM/DD hh:mm:ss), AP\_LIG\_FASE1 (kW), AP\_LIG\_FASE2 (kW) and AP\_LIG\_FASE3 (kW).
- Energy Analyzers—Power (5 min basis): timestamp (YYYY/MM/DD hh:mm:ss), AP\_POW\_FASE1 (kW), AP\_POW\_FASE2 (kW) and AP\_POW\_FASE3 (kW).
- Energy Efficiency Indicators—OCC, ORCC, SONCCIT, SOFFCCIT and NPSONC (10 min basis): timestamp (YYYY/MM/DD hh:mm:ss) and anomaly (true/false).
- Energy Efficiency Indicators—OCC (10 min basis): timestamp (YYYY/MM/DD hh:mm:ss) and anomaly (true/false).
- Energy Efficiency Indicators—ORCC (10 min basis): timestamp (YYYY/MM/DD hh:mm:ss) and anomaly (true/false).
- Energy Efficiency Indicators—SONCCIT (10 min basis): timestamp (YYYY/MM/DD hh:mm:ss) and anomaly (true/false).

- Energy Efficiency Indicators—SOFFCCIT (10 min basis): timestamp (YYYY/MM/DD hh:mm:ss) and anomaly (true/false).
- Energy Efficiency Indicators—NPSONC (10 min basis): timestamp (YYYY/MM/DD hh:mm:ss) and anomaly (true/false).

In this regard, the techniques used during the entire preprocessing stage were the following:

- Data cleaning: removing all missing and null values, as well as inconsistencies found in each data source;
- Data transformation: normalizing all data to the same period (10 min) to facilitate the aggregation of data and their analysis;
- Data reduction: simplifying the data with the aim of providing meaningful data.

Once the data were normalized and standardized, the data sample consisted of 5,088,765 records corresponding to EEI data and the set of variables of the BlueNet building. From all of these variables, only the most relevant ones were selected for use in the study. A list of these variables is summarized in Appendix A.

Initially, the 5,088,765 records of the dataset were analyzed. From these records, 636,848 missing data records were found, corresponding to null values. These records were filtered and removed from the data sample. After the exclusion of outliers, the next requirement was ensuring the data quality. Thus, a set of rules were established to guarantee the data quality:

- Detect every outlier in the data sample and fix it with the average value of data dispersion through DM techniques.
- Data consistency validation: every sample of data requires at least one value for each 30 min period.

As a consequence of this first phase of data quality control, 636,848 records were filtered in the preprocessing task, corresponding to null values. After applying basic rules to ensure data consistency, 34 records were filtered. Finally, the data sample was reduced from 5,088,765 records to 4,451,883 records.

Once the preprocessing phase was carried out, a set of 25 relevant variables was selected from the total of variables measured in the BlueNet building for the tasks of data classification and data prediction (described in Appendix A). This selection was carefully analyzed by studying the degree of the influence of each selected variable through Principal Component Analysis (PCA) techniques [73]. At first, the standardization of continuous variables was carried out, so each one contributed equally to the analysis. Second, a covariance matrix was established to identify correlations between the selected variables; pairs that had a positive score were correlated, and pairs that had a negative score were inversely correlated. Third, the eigenvectors and eigenvalues of the covariance matrix were computed to identify the principal components. These principal components are new variables constructed as linear combinations or mixtures of the initial variables; principal components are uncorrelated and provide the maximum information based on their variance. On the basis of the results, the most meaningful variables were selected, providing 25 relevant variables. Most of the variables analyzed were used in other similar research studies [74,75].

## 3.2. Classification Module

The classification module was developed through DM techniques with the aim of classifying days with similar features and comparing these days to quantify the EE. These clusters were developed with the aim of extracting insights about how different features affected days in EE terms. The second aim of these comparisons is to determine the clusters in which the detected anomalies are concentrated and infer key features that can provide some insights to reveal anomalies. Once the classification and detection are carried out, the main objective is to predict days with anomalies and then apply policies to improve the EE. This classification module uses the most influential information supplied by EEIs;

HVAC, lighting, indoor and outdoor environmental information; BlueNet occupation; and holiday/work schedule.

In the first instance, data for the classification module were analyzed through Generalized Rule Induction (GRI) [76] and the Apriori algorithm [77] to study the correlation between the selected variables and consumption. GRI begins with the original set *S* as a root node. It iterates through every unused attribute of the set *S* and calculates the entropy H(S) or the information gain IG(S) of that attribute. It selects the attribute with the smallest entropy (or largest IG), and the set *S* is partitioned by the selected attribute. Finally, the algorithm continues recursion on each subset.

Entropy H(S) is a measure of the amount of uncertainty in the dataset *S*:

$$H(S) = \sum_{x \in X} p(x) \log_2 p(X)$$

where *S* is the current dataset for which entropy is being calculated, *X* is the set of classes in *S*, and p(x) is the proportion of the number of elements in class *x* to the number of elements in set *S*.

Information Gain IG(A) is the measure of the difference in entropy from before to after the set *S* is split on the basis of an attribute *A*. In other words, it measures how much uncertainty in *S* was reduced after splitting set *S* on the basis of attribute *A*:

$$IG(S, A) = H(S) - \sum_{t \in T} p(t)H(t) = H(S) - H(S|A)$$

where H(S) is the entropy of set *S*; *T* is the subsets created by splitting the set *S* by attribute *A*, such as  $S = \bigcup_{t \in T} t$ ; p(t) is the proportion of the number of elements in *t* to the number of elements in set *S*; and H(t) is the entropy of subset *t*.

Additionally, the Apriori algorithm is used for frequent item set mining and association rule learning, and it uses breadth-first search and a hash tree structure (Algorithm 1).

```
Algorithm 1. A Priori with breadth-first search.
```

```
Apriori(T, \varepsilon)
             L_1 \leftarrow \{large \ 1 - itemsets\}
            k \leftarrow 2
             while L_{k-1} \neq \emptyset
                    C_k \longleftarrow \{a \cup \{b\} \mid a \in L_{k-1} \land b \notin a\} - \{c \mid \{s \mid s \subseteq c \land \mid s \mid = k-1\} \not\subseteq L_{k-1}\}
                    for transformations t \in T
                                    C_t \leftarrow \{c \mid c \in C_k \land c \subseteq t\}
                                    for candidates c \in C_t
                                         count[c] \leftarrow -count[c+1]
                      L_k \leftarrow \{c \mid c \in C_k \land count[c] \ge \varepsilon\}
                      k \leftarrow k + 1
             return \bigcup_k L_k
where
T—The set of data;
\epsilon—Confidence threshold;
k—Size of the set of candidate items;
C_k—Candidate set at level k;
c—Candidate c;
count[c]—Pointer to the candidate set c.
```

As a result, 24 features were selected from the initial set of 25 variables that were previously filtered. These features provided more detailed and useful information to develop the classification module through DM techniques with the aim of supplying the most accurate results possible. The features selected for the classification module are shown in Table 1.

Feature	Unit	_
AEMET_AT	°C	_
AP_COMPRESSOR	kW	
AP_COMPRESSOR_MINUTES	minutes	
AP_COMPRESSORS_MEAN	kW	
AP_LIGTH	kW	
DATE	DD/MM/YY hh:min:ss	
HOLIDAY	DD/MM/YY	
LIG_MINUTES	minutes	
NPSONC_MINUTES	minutes	
NPSONC_PERIODNUMBER	integer	
NPSONC_TIMES	integer	
OCC_STOPPOINTS	integer	
ORCC_PERIODNUMBER	integer	
ORCC_MINUTES	minutes	
ORCC_TIMES	integer	
PRESENCE_ID_ENT	DD/MM/YY hh:min:ss	
PRESENCE_ID_EXI	DD/MM/YY hh:min:ss	
SOFFCCIT_MINUTES	minutes	
SOFFCCIT_PERIODNUMBER	integer	
SOFFCCIT_TIMES	integer	
SONCCIT_AP	kŴ	
SONCCIT_MINUTES	minutes	
SONCCIT_PERIODNUMBER	integer	
SONCCIT_TIMES	integer	

Table 1. Selected features.

Subsequently, the resultant variables, also called features, were analyzed and classified according to outdoor temperature (AEMET\_AT), occupant presence (PRESENCE\_ID\_ENT and PRESENCE\_ID\_EXI) and HVAC compressor consumption (AP\_COMPRESSORS\_MEAN).

Firstly, a filter was applied to exclude all data without compressor measurements. On the one hand, after obtaining the sample without null compressor function values, an analysis of this sample was carried out to realize the segmentation of the data according to outdoor temperature (AEMET\_AT). As a result of applying a binning algorithm, 8 groups of outdoor temperature (TE\_Mean\_BIN) were defined with a range from 3.56 °C to 39.52 °C in 5 °C intervals, as is shown in Table 2. The binning algorithm is used to reduce the effects of minor observation errors, and it carried out bucketing, where bins have an equal frequency of 5 degrees Celsius following the formula:

$$L = \frac{\max(x) - \min(x)}{n} \tag{1}$$

where *L* is the length of the bucket, and *n* is the number of buckets.

Table 2.	External	temperature	binn	ing
				• • •

Bin	Lower	Upper
1	≥3,562,762	<8,562,762
2	≥8,562,762	<13,562,762
3	≥13,562,762	<18,562,762
4	≥18,562,762	<23,562,762
5	≥23,562,762	<28,562,762
6	≥28,562,762	<33,562,762
7	≥33,562,762	<38,562,762
8	≥38,562,762	<39,529,412

On the other hand, the distribution of the building occupancy was analyzed by applying statistical methods and defining 3 clear groups: a group with fewer than 50

persons, which mainly represents holidays; a group with between 50 and 110 persons, which indicates days with medium occupancy; and finally, a group with a range between 110 and 160 persons corresponding to working days with high occupancy, which contains the majority of the sample, as is shown in Figure 2.



Figure 2. Occupancy distribution.

Finally, the data were represented in terms of compressor consumption (AP\_COMPRESS ORS\_MEAN), external temperature (AEMET\_AT) and occupancy (PRESENCE\_ID\_ENT and PRESENCE\_ID\_EXI) to be further analyzed. Furthermore, a discriminant analysis between workdays and non-workdays was performed to provide more depth to our model. The sample distribution is illustrated in Figure 3.



Figure 3. Sample compressor consumption distribution.

After representing and studying the data distribution through the different features, a classification model based on a C&RT decision tree using Gini impurity measures was carried out, with the aim of modeling the sample based on the mean daily active power of the compressors (AP\_COMPRESSORS\_MEAN). As a result, a set of rules were obtained. C&RT is used for both classification and regression, and it uses the Gini Index (*GI*) criterion to split a node into subnodes. It starts with the training set as a root node, and after splitting the root node in two, it splits the subsets using the same logic recursively until it finds that further splits will not result in any pure subnodes or reaches the maximum number of leaves in a growing tree. The Gini Index is expressed as follows:

$$GI = \sum_{i=0}^{c} P_i (1 - P_i)$$

where *c* is total classes, and  $P_i$  is the probability of class *i*.

This set of rules was the basis for substantiating our clustering model. The scheme of this classification algorithm is shown in Figure 4.



**Figure 4.** Set of rules for the clustering of the sample with the objective of average active power compressor versus temperature and work or non-workdays.

As can be seen at the first level of the decision tree, the sample was classified into 2 groups distinguished by external temperature (*AEMET\_AT*). On the one hand, groups with lower external temperature ranging from  $3.5 \,^{\circ}$ C to  $23.5 \,^{\circ}$ C correspond to groups 1, 2, 3 and 4. On the other hand, groups with higher external temperature ranging from  $23.5 \,^{\circ}$ C to  $39.5 \,^{\circ}$ C correspond to groups 5, 6 and 7. Once this first classification was carried out, it was possible to observe a division between days with lower temperature: one group has external temperature ranges from  $3.5 \,^{\circ}$ C to  $18.5 \,^{\circ}$ C (1, 2 and 3) with very little compressor use, and another group has a comfortable temperature range ranging from  $18.5 \,^{\circ}$ C to  $23.5 \,^{\circ}$ C (4) with slight compressor use.

These 3 groups were derived from the following clusters analyzed to study EE in the BlueNet building.

#### 3.2.1. Cluster 1

In cluster 1 (Figure 5), it was possible to observe a classification of days with lower external temperature ranging from  $3.5 \,^{\circ}$ C to  $18.5 \,^{\circ}$ C. This group (158 days) had an average compressor consumption ranging from  $2.566 \,$ kW/h to  $2.583 \,$ kW/h on working days with an effect of -0.006 and 0.011, respectively. On non-working days (*LaborFestive* equals 1), the average consumption of the compressor ranged from  $2.067 \,$ kW/h to  $2.409 \,$ kW/h with an effect of  $-0.18 \,$  and 0.114, respectively, which indicates the effect on the entropy—average—when removing a value from the cluster. Furthermore, it should be noted that most of the days followed this distribution, while scattered days usually coincided with individual cases of high consumption over a short period of time. In addition, non-working days had a homogeneous distribution with compressors having quite low average active power consumption, very close to  $2 \,$ kW/h.



Figure 5. Graphical representation of distribution of cluster 1.

# 3.2.2. Cluster 2

In cluster 2 (Figure 6), it was possible to observe a classification of days with a comfortable external temperature that ranged between 18.5 °C and 23.5 °C. This group (38 days) had an average compressor consumption of 3.086 kW/h on workdays with an effect of -0.016. On non-working days, the compressor average was 2.934 kW/h with an effect of -0.136. In addition, it was possible to observe a wider dispersion and high similarity between workdays and non-workday cases. This is caused by a comfortable external temperature ranging between 18.5 °C and 23.5 °C. In addition, in Figure 6, it is possible to observe the number of minutes in which the HVAC system was switched on (*C\_FunctionMins*).



Figure 6. Graphical representation of distribution of cluster 2.

# 3.2.3. Cluster 3

After analyzing cluster 1 and cluster 2, cluster 3 was the most relevant and interesting, and it covered the majority of cases with high external temperature. This cluster, with 116 days, has a somewhat complex distribution because it was analyzed carefully and split into 3 well-defined groups, as is shown in Figure 7.



Figure 7. Graphical representation of distribution of cluster 3.

# Cluster 3.1

Cluster 3.1 contains the majority of non-working days, in which the compressor had a low range of average energy consumption, with values between 1.7 kW/h and 3 kW/h for the entire period covered in the data sample (26 days) (Figure 8).



Figure 8. Graphical representation of distribution of cluster 3.1.

## Cluster 3.2

Cluster 3.2 includes the majority of working days during the year 2011 until the months of July–August (22 days), during which consumption was higher, with values of 3 kW/h and 5 kW/h, except for a day on which data were dispersed to almost 8 kW/h. As it is possible to observe, this cluster corresponds to a season with more moderate temperatures, as is illustrated in Figure 9. Furthermore, in the illustration in Figure 9, it is possible to observe the number of minutes in which the HVAC system was switched on (*C\_FunctionMins*), providing further detail.



Figure 9. Graphical representation of distribution of cluster 3.2.

# Cluster 3.3

Finally, cluster 3.3 corresponds to most of the working days during 2012 from April to October. This cluster (68 days) is characterized by days with a large dispersion, caused by a hotter season with higher temperatures, and an average compressor consumption ranging between 3 kW/h and 7 kW/h in a regular manner, as is illustrated in Figure 10.



Figure 10. Graphical representation of distribution of cluster 3.3.

In this case, the behavior of all days is characterized by a high compressor consumption and a high number of changes in compressor consumption. Thus, on days with similar environmental characteristics, the features that indicate the grades of efficiency comprise lower compressor consumption, uniform compressor consumption and a small number of slight changes in compressor consumption.

## 3.3. Energy Efficiency Prediction Module

Once the developed classification module was implemented, the results of this module were carefully studied. As a result of the study, an energy efficiency prediction module was developed. The purpose of this module was mainly to obtain a metric of EE for this study. Therefore, clusters were studied independently, and the EE estimation was based on historical behavior inferred through statistical methods.

First, every cluster was selected, and a distribution analysis for each sample was carried out. The distribution analysis was based on an energy consumption histogram. This distribution curve follows a distribution that is similar to a Gaussian distribution; thus,



the EE categories were fitted with a function of the sample distribution, as is shown in Figure 11.

Figure 11. Gaussian distribution of energy consumption in cluster 3.3.

Thus, the days were categorized into 5 EE categories based on cluster dispersion. These categories were established on the basis of the normal distribution, which approximates a Gaussian distribution. The limits for each category comprise the following segments:  $\mu - 3/2 \sigma$ ,  $\mu - 1/2 \sigma$ ,  $\mu + 1/2 \sigma$  and  $\mu + 3/2 \sigma$ , where  $\mu$  is the mean, and  $\sigma$  is covariance. As a result, the consumption on each day was categorized according to every cluster characteristic following the classification based on the data distribution shown in Figure 12.



Figure 12. EE category prediction for cluster 3.3 and energy consumption behaviors.

Besides EE categories for each cluster, this module is able to provide some insights that can help to reveal and detect the patterns and anomalies observed on inefficient days. For example, the consumption behaviors on 4 October 2012 (cat. 1) and 20 September 2012

(cat. 5) had the same features (cluster 3.3), which comprise the same range of outdoor temperatures, number of persons and other environmental conditions, but the energy consumption behavior was very different, as it was highlighted through the EE category carried out by the EE prediction module. Figure 12 shows a wide range of consumption (0–6 kWh vs. 0–12 kWh), a greater number of stop points (5 OCC vs. 3 OCC) and regime changes (43 ORCC with 293 min of anomalous function vs. 77 ORCC with 532 min of anomalous function), a less softened energy consumption curve with a larger variation in energy consumption changes at the same time, etc. All of these results can indicate an anomaly, such as HVAC breakdown, poorly dimensioned HVAC, a possible isolation problem in the room and other issues.

In summary, the EE prediction module indicates the EE category for every day as a function of the historical behavior for the cluster sample. In addition, the EE prediction module, together with the classification module, can indicate possible anomalous consumption patterns that occur in the building and other possible anomalies. These cases were studied, and the results show a tight correlation between the detected OCC and ORCC anomalies and the index of EE, as shown in Tables 3–7. At the same time, it is observed that SOFFCCIT, SONCCIT and NPSONC anomalies are not correlated with inefficient consumption behavior.

**Table 3.** Cluster 1 anomaly distribution. Each EE Category is associated with a color as depicted in Figure 12.

Cluster 1	Days	OCC	ORCC	SOFFCCIT	SONCCIT	NPSONC	TOTAL
EE Categ. 1	0	-	-	-	-	-	0
EE Categ. 2	9	0	1	1	1	9	12
EE Categ. 3	8	0	5	9	6	8	28
EE Categ. 4	2	0	1	1	0	2	4
EE Categ. 5	0	0	1	0	0	0	1
	19	0	8	11	7	19	45

**Table 4.** Cluster 2 anomaly distribution. Each EE Category is associated with a color as depicted in Figure 12.

Cluster 1	Days	OCC	ORCC	SOFFCCIT	SONCCIT	NPSONC	TOTAL
EE Categ. 1	3	0	0	0	0	1	1
EE Categ. 2	8	0	0	0	0	1	1
EE Categ. 3	13	0	0	0	0	3	3
EE Categ. 4	12	0	2	0	0	0	2
EE Categ. 5	2	0	2	0	0	1	3
	38	0	4	0	0	6	10

**Table 5.** Cluster 3.1 anomaly distribution. Each EE Category is associated with a color as depicted in Figure 12.

Cluster 1	Days	OCC	ORCC	SOFFCCIT	SONCCIT	NPSONC	TOTAL
EE Categ. 1	0	-	-	-	-	-	0
EE Categ. 2	11	0	1	0	0	7	8
EE Categ. 3	8	0	1	0	0	4	5
EE Categ. 4	4	0	0	0	0	2	2
EE Categ. 5	3	0	0	0	0	1	1
	26	0	2	0	0	14	16

Cluster 1	Days	OCC	ORCC	SOFFCCIT	SONCCIT	NPSONC	TOTAL
EE Categ. 1	0	-	-	-	-	-	0
EE Categ. 2	7	0	3	0	0	3	6
EE Categ. 3	12	0	3	0	0	3	6
EE Categ. 4	2	0	0	0	0	0	0
EE Categ. 5	1	0	1	0	0	0	1
	22	0	7	0	0	6	13

**Table 6.** Cluster 3.2 anomaly distribution. Each EE Category is associated with a color as depicted in Figure 12.

**Table 7.** Cluster 3.3 anomaly distribution. Each EE Category is associated with a color as depicted in Figure 12.

Cluster 1	Days	OCC	ORCC	SOFFCCIT	SONCCIT	NPSONC	TOTAL
EE Categ. 1	5	0	0	0	0	0	0
EE Categ. 2	16	8	12	0	0	1	21
EE Categ. 3	20	13	16	0	0	1	30
EE Categ. 4	18	14	14	0	0	2	30
EE Categ. 5	5	4	5	0	0	0	9
	64	39	47	0	0	4	90

The above-mentioned correlation between detected OCC and ORCC anomalies versus the index of EE highlights not only the number of anomalies but also the period (in minutes) of each anomaly for ORCC and the number of stop points for OCC, as shown in Table 8. This insight is even more relevant when one of the principles of EE consists in softening the consumption curve, aiming to avoid peaks and large changes in consumption. Furthermore, OCC and ORCC anomalies denote large changes and peaks in consumption, which means that the HVAC system is poorly dimensioned or calibrated, the HVAC system is too powerful, or a possible breakdown in a temperature sensor has occurred. On the other hand, a compressor with high OCC is prone to break and to have a shortened lifetime. Table 8 highlights the correlation between an increasing number of stop points for OCC and an increasing number of minutes with ORCC in each period when EE decreases. Additionally, in Table 9 are detailed the average of OCC and ORCC anomalies per day and its correlation with each EE Category.

**Table 8.** Stop points and minutes for OCC and ORCC anomalies vs. EE category. Each EE Category is associated with a color as depicted in Figure 12.

Cluster	Anomaly	Days	Categ. 1	Categ. 2	Categ. 3	Categ. 4	Categ. 5
Cluster 1	OCC	5	0	0	0	0	0
Cimbici 1	ORCC	16	0	313 (1)	1769 (5)	418 (1)	544 (1)
Cluster 2	OCC	20	0	0	0	0	0
Chubici 2	ORCC	18	0	0	0	724 (2)	658 (2)
Cluster 2 1	OCC	20	0	0	0	0	0
Chubici 0.1	ORCC	18	0	340 (1)	315 (1)	0	0
Cluster 3.2	OCC	20	0	0	0	0	0
Chubici 0.2	ORCC	18	0	1020 (3)	945 (3)	0	543 (1)
Cluster 3 3	OCC	20	0	64 (8)	142 (13)	158 (14)	39 (4)
Ciusier 5.5 –	ORCC	18	0	5026 (12)	7278 (16)	5881 (14)	2671 (5)
Total		64	0	64/6699	142/10307	158/7023	39/4416

Cluster	Anomaly	Days	Categ. 1	Categ. 2	Categ. 3	Categ. 4	Categ. 5
Cluster 1	OCC	5	0	0	0	0	0
Chubici 1	ORCC	16	0	313	353.8	418	544
Cluster ?	OCC	20	0	0	0	0	0
Cluster 2	ORCC	18	0	0	0	362	329
Cluster 3.1	OCC	20	0	0	0	0	0
Cluster 5.1	ORCC	18	0	340	315	0	0
Cluster 3.2	OCC	20	0	0	0	0	0
Cluster 5.2	ORCC	18	0	462.7	437.7	0	543
Cluster 3 3	OCC	20	0	8	10.9	11.3	9.8
Cluster 5.5 =	ORCC	18	0	418.8	454.9	420.1	534.2
Total		64	0	8/394.1	10.9/412.3	11.3/413.1	9.8/490.7

**Table 9.** Average stop points and minutes per day for OCC and ORCC anomalies. Each EE Category is associated with a color as depicted in Figure 12.

## 4. Summary and Experimental Results

The objective of the present work was to present a decision support system to optimize Energy Management and Control Systems (EMCSs) in smart buildings for any energyconsuming operations. Thus, in order to achieve the main purpose of this work, a model to extract knowledge to realize the efficient management of energy consumption was developed. This model was based on a classification module and a prediction module developed through DM techniques and statistical inference. All BlueNet variables were considered for this EE classification study (indoor, outdoor, energy analyzers, EEIs and work schedule). In addition, the knowledge extracted from data and EE experts was considered for this purpose. Once the study was carried out, five clusters were identified. Each cluster was defined by the following set of conditions:

- Cluster 1: This cluster represents days with lower external temperature ranging from 3.5 °C to 23.5 °C. This cluster does not discriminate between energy consumption or between work and non-workdays.
- Cluster 2: This cluster represents days with a softer curve of intermediate temperature ranging from 23.5 °C to 28.5 °C. This cluster does not discriminate between energy consumption or between work and non-workdays.
- Cluster 3.1: This cluster groups days with higher external temperature ranging from 28.5 °C to 39.6 °C, in which most of the days are non-working days with low energy consumption.
- Cluster 3.2: This cluster represents days with higher external temperature ranging mainly from 28.5 °C to 33.5 °C. Most of the selected days are working days with high energy consumption.
- Cluster 3.3: This cluster groups days with higher external temperature ranging mainly from 28.5 °C to 39.6 °C, in which most of the selected days are working days with high energy consumption.

Once clusters were defined, days that belong to the same cluster were compared to observe differences among behaviors and, consequently, the effects of the different variables on each day. In order to validate and quantify the results, the results were analyzed and corroborated by EE experts of the BlueNet building, and additionally, a module for EE prediction was developed. The aim of the EE prediction module was to evaluate EE according to the type of cluster analyzed. The results obtained were applied to the sample of 509 days with the following results.

All of the results in this study were corroborated by EE experts. In summary, only 3.25% of the total days classified were considered to be very efficient, 29% were efficient and

41.36% were classified with normal EE. The remaining 20.29% of the days were considered to be energy inefficient, and 6.36% of days were classified as very inefficient, indicating a large margin for EE improvement, as detailed in Table 10.

Moreover, in this study, days that represent anomalous behavior were detected with the support of the classification module and prediction module. These anomalous behaviors usually correspond to days that were very inefficient. In addition, these anomalous behaviors can indicate possible problems that are affecting the BlueNet building (HVAC breakdown, incorrectly dimensioned and other problems), as shown in Tables 2–8. With the results in this study, system integration with EE software could yield substantial benefits: detection of anomalous behaviors in energy consumption, facilitation of energy savings, large profits due to consumption reduction and environmentally friendly management of the building.

Cluster Days Categ. 1 Categ. 2 Categ. 3 Categ. 4 Categ. 5 Cluster 1 32.35% 5.25% 1710% 48.87% 13.53% Cluster 2 91 7.89% 21.05% 34.21% 31.58% 5.27% Cluster 3.1 26 0% 42.31% 30.77% 15.39% 11.53% Cluster 3.2 22 0% 54.55% 9.09% 4.54% 31.82% 7.81% Cluster 3.3 64 7.81% 25% 31.25% 28.13% Total 374 3.25% 29% 41.36% 20.29% 6.36%

**Table 10.** Distribution of days for each cluster based on EE category. Each EE Category is associated with a color as depicted in Figure 12.

## 5. Conclusions

After a bibliographical review, an intensive field of research with strong interest in EE was identified. This review revealed that similar approaches have been considered, but no other works have been performed with the aim of rendering support in EE classification and prediction. Specifically, systems able to measure EE and support the detection of anomalies have not been investigated with similar approaches or results.

This paper presents a system for the optimization of energy consumption in smart buildings. The main purposes of this work are the following: extract the knowledge hidden in building data, develop a classification module and build an EE prediction module that helps to predict EE for each day. This proposal was tested in the BlueNet building scenario, but it is applicable to any other building and is not specific to BlueNet.

The process of energy optimization was carried out through a hybrid system whose cornerstone is a classification module and an EE prediction module. The classification module is made up of a hybrid system based on a CR&T decision tree and a clustering model. As a result, this module is able to compare EE on days with similar characteristics, regardless of the season and other factors, which are difficult to compare without this information, and it is also able to highlight the presence of anomalous energy consumptions and other possible problems. This module provides an objective point of view that is key to measuring, comparing and predicting the EE for each day.

The EE prediction module is able to quantify the energy efficiency of each day, comparing days with similar EE conditions supported by previous cluster classifications. This module is able to measure and predict the efficiency for each day based on the knowledge extracted from historical data by applying statistical analysis. Furthermore, this module is able to unveil insights that highlight correlations between inefficiencies and anomalous behaviors.

Finally, this work presents a classification for each day of the historical data and their respective EE categories. These results were compared and corroborated by experts to, first, understand how the energy consumption is behaving and, second, understand the reasons for this behavior and how to enhance the efficiency. Furthermore, our approach highlights some evidence that days with less efficiency (6.36% and 20.29% of days respectively)

contain more anomalies and that these anomalies also occurred over a greater amount of time. Furthermore, this work unveils an exponential correlation between OCC and ORCC anomalies and EE.

Based on the results obtained in the BlueNet building and aligned with this research area, interesting future research lines that could be explored include the automation of the full process through AI techniques that are able to classify and predict EE in an analytical manner in real time and, based on anomalous behavior, determine how to apply actionable measures to correct and improve EE in the building automatically. On the other hand, this work aimed to support the gap in the understanding of building behavior, as cited in [39]. A future goal is the integration of this module with the commercial EE software called Eugene, whose owner is the company Isotrol.

Author Contributions: Conceptualization, M.P., F.B., E.P. and C.L.; methodology, M.P., F.B. and C.L.; software, M.P.; validation, M.P.; formal analysis, M.P. and F.B.; investigation, M.P., F.B., E.P. and C.L.; resources, M.P., F.B. and C.L.; data curation, M.P.; writing—original draft preparation, M.P.; writing—review and editing, M.P., F.B., E.P. and C.L.; visualization, M.P. and F.B.; supervision, M.P., F.B., E.P. and C.L.; funding acquisition, C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the European Commission—Framework Program 7, FP7-285229 KnoHolEM.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** All data and code will be made available on request to the correspondent author's email with appropriate justification.

**Acknowledgments:** The authors thank FP7-285229 KnoHolEM for providing funds for this project. The authors thank Isotrol for their collaboration in this work by providing knowledge and cooperation to extract the data from the BlueNet building database.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### Appendix A. Main Variables for Optimization in BlueNet Building

AEMET\_AT (°C): Outdoor air temperature in Celsius degrees outside of the BlueNet building.

AP\_AIR\_MIXERS (kW): Active power in air mixers of the HVAC system.

AP\_COMPRESSOR\_MEAN (kW): Average daily active power of the HVAC compressor. AP\_COMPRESSOR\_MINUTES (min): Total number of daily minutes of HVAC compressor use.

AP\_COMPRESSOR (kW): Active power of the HVAC compressor.

AP\_SPLITS (kW): Active power in splits of the HVAC system.

AP\_LIGTH (kW): Active power in the lighting system.

LABOR/FESTIVE (time): Date of a non-working day.

LIG\_MINUTES (min): Time in minutes in which lighting was working during a day.

LIG\_PERIODNUMBER: Total number of periods of lighting during a day.

NPSONC\_MINUTES (min): Time in minutes in which lighting was working and an NPSONC anomaly was detected.

NPSONC\_PERIODNUMBER: Total number of periods with an NPSONC anomaly during a day. A period constitutes the number of consecutive points with NPSONC anomalies detected.

NPSONC\_TIMES: Number of times with an NPSONC anomaly during a day.

OCC\_STOPPOINTS: Number of daily On-Off operations in HVAC compressor (OCC).

ORCC\_PERIODNUMER: Total number of periods with an operational regime change (ORCC) in HVAC compressor during a day.

ORCC\_MINUTES (min): Time in minutes in which an operational regime change in HVAC compressor (ORCC) anomaly was detected.

ORCC \_ TIMES: Number of times with an operational regime change in HVAC compressor (ORCC) anomaly during a day.

SOFFCCIT \_MINUTES (min): Time in minutes in which the HVAC compressor was switched off and an abnormal change in interior temperature (SOFFCCIT) was detected.

SOFFCCIT\_PERIODNUMBER: Total number of periods in which the HVAC compressor was switched off and an abnormal change in interior temperature (SOFFCCIT) is detected during a day.

SOFFCCIT\_TIMES: Number of times in which the HVAC compressor was switched off and an abnormal change in interior temperature (SOFFCCIT) was detected during a day.

SONCCIT\_AP (kW): Active power waste in the HVAC system while the HVAC compressor was switched on and an abnormal change in interior temperature (SONCCIT) was detected.

SONCCIT\_MINUTES (min): Time in minutes in which the HVAC compressor was switched off and an abnormal change in interior temperature (SONCCIT) was detected.

SONCCIT\_PERIODNUMBER: Total number of periods in which the HVAC compressor was switched off and an abnormal change in interior temperature (SONCCIT) was detected during a day.

SONCCIT\_TIMES: Number of times in which the HVAC compressor was switched off and an abnormal change in interior temperature (SONCCIT) was detected during a day.

ZIGBEE\_AT (°C): Indoor air temperature in Celsius degrees inside the BlueNet building.

## References

- Rashid, Y.R.; Sulaiman, M.S.; Aziz, A.; Selamat, H.; Yani, A.H.M.; Kandar, M.Z. Greening government's office buildings: PWD Malaysia experiences. *Procedia Eng.* 2011, 21, 1056–1060. [CrossRef]
- Pérez-Lombard, L.; Ortiz, J.; Pout, C. A review on buildings energy consumption information. *Energy Build*. 2008, 40, 394–398. [CrossRef]
- Peters, G.P.; Andrew, R.M.; Canadell, J.G.; Fuss, S.; Jackson, R.B.; Korsbakken, J.I.; Le Quéré, C.; Nakicenovic, N. Key indicators to track current progress and future ambition of the Paris Agreement. *Nat. Clim. Chang.* 2017, 7, 118. [CrossRef]
- 4. EIA, U.S. *Key Energy Statistics*; Energy Information Administration: Washington, DC, USA, 2017.
- 5. EIA, U.S. Annual Energy Outlook; Energy Information Administration: Washington, DC, USA, 2018.
- Day, A.; Jones, P.; Turton, J. Development of a UK Centre for Efficient and Renewable Energy in Buildings (CEREB). *Renew. Energy* 2012, 49, 166–170. [CrossRef]
- de Alegría Mancisidor, I.; de Basurto Uraga, P.D.; de Arbulo López, P.R. European Union's renewable energy sources and energy efficiency policy review: The Spanish perspective. *Renew. Sustain. Energy Rev.* 2009, 13, 100–114. [CrossRef]
- 8. EEA. Financial Support for Energy Efficiency in Buildings 2018; EEA: Copenhagen, Denmark, 2018.
- 9. DOE/EIA. Annual Energy Outlook 2021; DOE/EIA: Washington, DC, USA, 2021.
- 10. Yamagata, Y.; Seya, H. Simulating a future smart city: An integrated land use-energy model. *Appl. Energy* **2013**, *112*, 1466–1474. [CrossRef]
- 11. Lazaroiu, G.C.; Roscia, M. Definition methodology for the smart cities model. Energy 2012, 47, 326–332. [CrossRef]
- Personal, E.; Guerrero, J.I.; Garcia, A.; Peña, M.; Leon, C. Key performance indicators: A useful tool to assess Smart Grid goals. Energy 2014, 76, 976–988. [CrossRef]
- Vadiee, A.; Martin, V. Thermal energy storage strategies for effective closed greenhouse design. *Appl. Energy* 2013, 109, 337–343.
   [CrossRef]
- 14. Singh, R.; Tiwari, G. Energy conservation in the greenhouse system: A steady state analysis. Energy 2010, 35, 2367–2373. [CrossRef]
- 15. Djevic, M.; Dimitrijevic, A. Energy consumption for different greenhouse constructions. *Energy* 2009, 34, 1325–1331. [CrossRef]
- 16. Hoseini, A.G.; Dahlan, N.D.; Berardi, U.; Hoseini, A.G.; Makaremi, N.; Hoseini, M.G. Sustainable energy performances of green buildings: A review of current theories, implementations and challenges. *Renew. Sustain. Energy Rev.* **2013**, 25, 1–17. [CrossRef]
- 17. Hoseini, A.G.; Dahlan, N.; Berardi, U.; Hoseini, A.G.; Makaremi, N. The essence of future smart houses: From embedding ICT to adapting to sustainability principles. *Renew. Sustain. Energy Rev.* **2013**, *24*, 593–607. [CrossRef]
- Dounis, A.I.; Caraiscos, C. Advanced control systems engineering for energy and comfort management in a building environment— A review. *Renew. Sustain. Energy Rev.* 2009, 13, 1246–1261. [CrossRef]
- 19. Omer, A.M. Renewable building energy systems and passive human comfort solutions. *Renew. Sustain. Energy Rev.* 2008, 12, 1562–1587. [CrossRef]
- 20. Xia, C.; Zhu, Y.; Lin, B. Renewable energy utilization evaluation method in green buildings. *Renew. Energy* **2008**, *33*, 883–886. [CrossRef]

- 21. Peña, M.; Biscarri, F.; Guerrero, J.I.; Monedero, I.; León, C. Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. *Expert Syst. Appl.* **2016**, *56*, 242–255. [CrossRef]
- 22. Zanella, A.; Bui, N.; Castellani, A.; Vangelista, L.; Zorzi, M. Internet of Things for Smart Cities. *IEEE Int. Things J.* 2014, 1, 22–32. [CrossRef]
- 23. Singh, D.; Tripathi, G.; Jara, A.J. A survey of Internet-of-Things: Future vision, architecture, challenges and services. *Int. Things* (*WF-IoT*) **2014**, *1*, 287–292. [CrossRef]
- 24. Al-Fuqaha, A.; Guizani, M.; Mohammadi, M.; Aledhari, M.; Ayyash, M. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 2347–2376. [CrossRef]
- 25. Rathore, M.M.; Paul, A.; Hong, W.-H.; Seo, H.; Awan, I.; Saeed, S. Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data. *Sustain. Cities Soc.* **2018**, *40*, 600–610. [CrossRef]
- 26. Pawar, P.; Kumar, M.T.; Vittal, K.P. An IoT based Intelligent Smart Energy Management System with accurate forecasting and load strategy for renewable generation. *Measurement* **2020**, *152*, 107187. [CrossRef]
- 27. Palattella, M.; Dohler, M.; Grieco, A.; Rizzo, J.; Torsner, T.; Engel, L.; Ladid, L. Internet of Things in the 5G era: Enablers architecture and business models. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 510–527. [CrossRef]
- Skouby, K.E.; Lynggaard, P. Smart home and smart city solutions enabled by 5G, IoT, AAI, and CoT services. In Proceedings of the 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysuru, India, 27–29 November 2014; pp. 874–878.
- Goraczek, M.; Pereira, G.; Falco, E.; Kleinhans, R.; Parycek, P. Using Fuzzy Cognitive Maps as Decision Support Tool for Smart Cities. In CeDEM Asia 2016: Proceedings of the International Conference for E-Democracy and Open Government, Asia 2016, 7–9 Decemer 2016, Deagu, Korea; Edition Donau-Universität Krems: Krems an der Donau, Austria, 2017.
- Somu, N.; Raman, G.; Ramamritham, K. A hybrid model for building energy consumption forecasting using long short term memory networks. *Appl. Energy* 2020, 261, 114131. [CrossRef]
- Ikhide, M.; Egaji, A.; Ahmed, A. Developing Energy Control and Optimisation Methodology for Built Environment of the Future. *Renew. Energy Sustain. Build.* 2020, 45, 567–577. [CrossRef]
- Fan, C.; Sun, Y.; Xiao, F.; Ma, J.; Lee, D.; Wang, J.; Tseng, Y.C. Statistical investigations of transfer learning-based methodology for short-term building energy predictions. *Appl. Energy* 2020, 262, 114499. [CrossRef]
- Hosseinnezhad, V.; Shafie-Khah, M.; Siano, P.; Catalao, J.P.S. An Optimal Home Energy Management Paradigm with an Adaptive Neuro-Fuzzy Regulation. *IEEE Access* 2020, *8*, 19614–19628. [CrossRef]
- 34. Broujeny, R.S.; Madani, K.; Chebira, A.; Amarger, V.; Hurtard, L. Data-Driven Living Spaces' Heating Dynamics Modeling in Smart Buildings using Machine Learning-Based Identification. *Sensors* **2020**, *20*, 1071. [CrossRef]
- Pallonetto, F.; De Rosa, M.; Milano, F.; Finn, D.P. Demand response algorithms for smart-grid ready residential buildings using machine learning models. *Appl. Energy* 2019, 239, 1265–1282. [CrossRef]
- Fan, C.; Xiao, F.; Yan, C.; Liu, C.; Li, Z.; Wang, J. A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. *Appl. Energy* 2018, 235, 1551–1560. [CrossRef]
- Reynolds, J.; Ahmad, M.W.; Rezgui, Y.; Hippolyte, J.L. Operational supply and demand optimisation of a multi-vector district energy system using artificial neural networks and a genetic algorithm. *Appl. Energy* 2019, 235, 699–713. [CrossRef]
- Olszewski, R.; Pałka, P.; Turek, A. Solving "Smart City" Transport Problems by Designing Carpooling Gamification Schemes with Multi-Agent Systems: The Case of the So-Called "Mordor of Warsaw". Sensors 2018, 18, 141. [CrossRef] [PubMed]
- Yoshino, H.; Hong, T.; Nord, N. IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods. *Energy Build.* 2017, 152, 124–136. [CrossRef]
- 40. Noailly, J. Improving the energy efficiency of buildings: The impact of environmental policy on technological innovation. *Energy Econ.* **2012**, *34*, 795–806. [CrossRef]
- Andrews, C.J.; Krogmann, U. Explaining the adoption of energy-efficient technologies in U.S. commercial buildings. *Energy Build*. 2009, 41, 287–294. [CrossRef]
- Colmenar-Santos, A.; de Lober, L.N.T.; Borge-Diez, D.; Castro-Gil, M. Solutions to reduce energy consumption in the man-agement of large buildings. *Energy Build.* 2013, 56, 66–77. [CrossRef]
- 43. Biyik, E.; Kahraman, A. A predictive control strategy for optimal management of peak load, thermal comfort, energy storage and renewables in multi-zone buildings. *J. Build. Eng.* **2019**, *25*, 100826. [CrossRef]
- Chumnanvanichkul, P.; Chirapongsananurak, P.; Hoonchareon, N. Three-level Classification of Air Conditioning Energy Consumption for Building Energy Management System Using Data Mining Techniques. In Proceedings of the 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), Bangkok, Thailand, 20 March–23 March 2019; pp. 611–615. [CrossRef]
- 45. Marinakis, V.; Karakosta, C.; Doukas, H.; Androulaki, S.; Psarras, J. A building automation and control tool for remote and real time monitoring of energy consumption. *Sustain. Cities Soc.* **2013**, *6*, 11–15. [CrossRef]
- Azadeh, A.; Saberi, M.; Ghaderi, S.; Gitiforouz, A.; Ebrahimipour, V. Improved estimation of electricity demand function by integration of fuzzy system and data mining approach. *Energy Convers. Manag.* 2008, 49, 2165–2177. [CrossRef]
- 47. Altwies, J.E.; Nemet, G.F. Innovation in the U.S. building sector: An assessment of patent citations in building energy control technology. *Energy Pol.* **2013**, *52*, 819–831. [CrossRef]

- 48. Pang, X.; Wetter, M.; Bhattacharya, P.; Haves, P. A framework for simulation-based real-time whole building performance assessment. *Build. Environ.* **2012**, *54*, 100–108. [CrossRef]
- Kolokotsa, D.; Kalaitzakis, K.; Antonidakis, E.; Stavrakakis, G. Interconnecting smart card system with PLC controller in a local operating network to form a distributed energy management and control system for buildings. *Energy Convers. Manag.* 2002, 43, 119–134. [CrossRef]
- 50. Nguyen, T.A.; Aiello, M. Energy intelligent buildings based on user activity: A survey. Energy Build. 2013, 56, 244–257. [CrossRef]
- 51. Escrivá-Escrivá, G. Basic actions to improve energy efficiency in commercial buildings in operation. *Energy Build.* **2011**, 43, 3106–3111. [CrossRef]
- 52. D'Agostino, D.; Cuniberti, B.; Bertoldi, P. Energy consumption and efficiency technology measures in European non-residential buildings. *Energy Build.* **2017**, *153*, 72–86. [CrossRef]
- 53. Koezjakov, A.; Urge-Vorsatz, D.; Crijns-Graus, W.; van den Broek, M. The relationship between operational energy demand and embodied energy in Dutch residential buildings. *Energy Build.* **2018**, *165*, 233–245. [CrossRef]
- Figueiredo, J.; Sá da Costa, J. A SCADA system for energy management in intelligent buildings. *Energy Build.* 2012, 49, 85–98.
   [CrossRef]
- Diakaki, C.; Grigoroudis, E.; Kolokotsa, D. Towards a multi-objective optimization approach for improving energy efficiency in buildings. *Energy Build.* 2008, 40, 1747–1754. [CrossRef]
- Diakaki, C.; Grigoroudis, E.; Kabelis, N.; Kolokotsa, D.; Kalaitzakis, K.; Stavrakakis, G. A multi-objective decision model for the improvement of energy efficiency in buildings. *Energy* 2010, 35, 5483–5496. [CrossRef]
- 57. Wu, W.; Guo, J.; Li, J.; Hou, H.; Meng, Q.; Wang, W. A multi-objective optimization design method in zero energy building study: A case study concerning small mass buildings in cold district of China. *Energy Build.* **2018**, *158*, 1613–1624. [CrossRef]
- 58. Doukas, H.; Nychtis, C.; Psarras, J. Assessing energy-saving measures in buildings through an intelligent decision support model. *Build. Environ.* **2009**, 44, 290–298. [CrossRef]
- Yang, R.; Wang, L. Development of multi-agent system for building energy and comfort management based on occupant behaviors. *Energy Build.* 2013, 56, 1–7. [CrossRef]
- 60. Klein, L.; Kwak, J.-Y.; Kavulya, G.; Jazizadeh, F.; Becerik-Gerber, B.; Varakantham, P.; Tambe, M. Coordinating occupant behavior for building energy and comfort management using multi-agent systems. *Autom. Constr.* **2012**, *22*, 525–536. [CrossRef]
- 61. Oldewurtel, F.; Parisio, A.; Jones, C.; Gyalistras, D.; Gwerder, M.; Stauch, V.; Lehmann, B.; Morari, M. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy Build.* **2012**, *45*, 15–27. [CrossRef]
- 62. Stamatescu, G.; Stamatescu, I.; Arghira, N.; Drăgana, C.; Făgărăşan, I. Data-driven methods for smart building AHU subsystem modelling. In Proceedings of the 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Bucharest, Romania, 21–23 September 2017; pp. 617–621.
- 63. Kim, H.; Stumpf, A.; Kim, W. Analysis of an energy efficient building design through data mining approach. *Autom. Constr.* **2011**, 20, 37–43. [CrossRef]
- 64. Petrushevski, F.; Gaida, S.; Beigelböck, B.; Sipetic, M.; Zucker, G.; Schiefer, C.; Schachinger, D.; Kastner, W. Semantic Building Systems Modeling for Advanced Data Analytics for Energy Efficiency. In Proceedings of the 15th IBPSA Conference, San Francisco, CA, USA, 7–9 August 2017; pp. 622–627.
- Schachinger, D.; Gaida, S.; Kastner, W.; Petrushevski, F.; Reinthaler, C.; Sipetic, M.; Zucker, G. An advanced data analytics framework for energy efficiency in buildings. In Proceedings of the 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), Berlin, Germany, 6–9 September 2016; pp. 1–4. [CrossRef]
- 66. Fan, C.; Xiao, F.; Li, Z.; Wang, J. Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. *Energy Build*. **2018**, *159*, 296–308. [CrossRef]
- Li, X.; Bowers, C.; Schnier, T. Classification of Energy Consumption in Buildings with Outlier Detection. *IEEE Trans. Ind. Electron.* 2009, 57, 3639–3644. [CrossRef]
- 68. Cardillo, E.; Li, C.; Caddemi, A. Embedded heating, ventilation, and air-conditioning control systems: From traditional technologies toward radar advanced sensing. *Rev. Sci. Instrum.* **2021**, *92*, 061501. [CrossRef]
- 69. Labeodan, T.; Zeiler, W.; Boxem, G.; Zhao, Y. Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation. *Energy Build.* **2015**, *93*, 303–314. [CrossRef]
- 70. Dhurandhar, A.; Dobra, A. Probabilistic Characterization of Random Decision Trees. J. Mach. Learn. Res. 2008, 9, 2321–2348.
- 71. Maimon, O.; Rokach, L. Data Mining and Knowledge Discovery Handbook, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2010.
- 72. Han, J.; Kamber, M. Data Mining. Concepts and Techniques, 1st ed.; Morgan Kaufmann: Burlington, MA, USA, 2001.
- 73. Sehgal, S.; Singh, H.; Agarwal, M.; Bhasker, V. Data analysis using principal component analysis 2014. In Proceedings of the 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), Greater Noida, India, 7–8 November 2014; pp. 45–48.
- 74. Hou, Z.; Lian, Z.; Yao, Y.; Yuan, X. Data mining-based sensor fault diagnosis and validation for building air conditioning system. *Energy Conv. Manag.* **2006**, *47*, 2479–2490. [CrossRef]
- Wu, S.; Clements-Croome, D. Understanding the indoor environment through mining sensory data—A case study. *Energy Build*. 2007, 39, 1183–1191. [CrossRef]

- 76. Zhang, C.; Zhang, S. Association Rule Mining: Models and Algorithms; Springer: Berlin/Heidelberg, Germany, 2002; ISBN 3-540-43533-6.
- 77. Dongre, J.; Prajapati, G.L.; Tokekar, S.V. The role of Apriori algorithm for finding the association rules in Data mining. In Proceedings of the 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques, Ghaziabad, India, 7–8 February 2014; pp. 657–660. [CrossRef]